Psychological Bulletin

165

James Deese, Editor
University of Virginia

Editorial Consultants

Norman H. Abrahams
George W. Albee
Chris Argyris
Justin Aronfreed
Helen Astin
Albert Bandura
William P. Banks
C. J. Bartlett
Harold P. Bechtoldt
Lewis Berger
Alvin S. Bernstein
Douglas A. Bernstein
James Bieri
Elliot M. Blass
Paul Bloomers
Robert C. Bolles
Gordon Bower
Joseph V. Brady
Martin D. S. Braine
Dana Bramel
Arthur H. Brayfield
George R. Breed
Louis Breger
Jack W. Brehm
Judson S. Brown
B. R. Bugelski
Arnold H. Buss
David P. Campbell
Donald T. Campbell
Douglas K. Candland
Virgil R. Carlson
Peter L. Carlton
John B. Carroll
Dorwin Cartwright
James V. Clark
Michael Cole
Andrew L. Comrey
Anne P. Constantinople
Stanley Coopersmith
Emory L. Cowen
Lee J. Cronbach
Robert G. Crowder
Gerald C. Davison
Robyn M. Dawes
Richard Degerman
Victor Denenberg
Gordon F. Derner
Ward Edwards
Dorothy H. Eichorn
Bruce K. Ekstrand
David Elkind
Doris R. Entwisle
Charles W. Eriksen
Shelby H. Evans
Ralph Exline
Seymour Feshbach
Samuel Fillenbaum
Ronald L. Flaugher
Cyril M. Franks
Bernard Z. Friedlander
Frank Friedlander
Herbert Friedman
Phillip S. Gallo
Gordon G. Gallup
Kenneth Gamble
P. James Geiwitz
Edwin E. Ghiselli
Gene V. Glass
Goldine C. Gleser
Lewis R. Goldberg
Lewis Gollub
Leo A. Goodman
Harrison Gough
Francis K. Graham
Bert Green
John T. Guthrie
Ralph Hakstian
Helen H. Hamilton
Willard W. Hartup
Norman Henderson
E. Mavis Hetherington
Philip Himelsteim
Julian Hochberg
Larry Hochhaus
Paul J. Hoffman
Robert A. Hogan
Peter Houri
Barrie E. Houston
Stewart H. Hulse
Raymond G. Hunt
Robert L. Isaacson
Ira Iscoe
Marvin Iverson
Margaret Ives
Murray E. Jarvik
Edward E. Jones
Lawrence Karlin
Richard E. Keesey
Herbert C. Kelman
Howard H. Kendler
Daniel P. Kimble
Herbert D. Kimmel
Walter Kintsch
David H. Krantz
John I. Lacey
Edward E. Lawler
Arnold A. Lazarus
Richard S. Lazarus
Russell C. Leaf
Edwin E. Locke
Vincent M. LoLordo
Joel F. Lubar
David T. Lykken
David Margules
Albert R. Marston
William M. McGlothlin
Joseph E. McGrath
Paul E. Meehl
Kenneth E. Moyer
James S. Myer
Sidney Newman
John R. Nichols
Donald A. Norman
David S. Olton
Martin T. Orne
Allan Parducci
Gordon L. Paul
Robert W. Payne
Dean Peabody
Robert F. Peck
Henry S. Pennypacker
Robert Perloff
E. Jerry Phares
Francis J. Pilgrim
Benjamin Pope
Andrew C. Porter
Leo J. Postman
K. Edward Renner
Robert A. Rescorla
Leonard G. Rorer
Milton J. Rosenberg
Leonard E. Ross
Haldor E. Rosvold
Julian B. Rotter
Aryeh Routtenberg
Nevitt R. Sanford
David R. Saunders
Richard A. Schmidt
D. L. Schurman
John Paul Scott
John W. Senders
Marvin E. Shaw
Donald L. Sherman
Lester C. Shine
Ronald E. Shor
Alexander W. Siegel
Jerome E. Singer
Norman Slamecka
Paul Slovic
M. Brewster Smith
Frank Sodetz
Robert Sommer
Julian C. Stanley
John A. Stern
Ezra Stotland
Percy H. Tannenbaum
Kenneth Terhune
Ethel Tobach
Jay A. Trowill
Tadeo Uno
Graham H. Vaughan
William Vogel
Roger E. Vogeler
Allan R. Wagner
S. Rains Wallace
William P. Wallace
Nancy Waugh
Roger Webb
Morton Weiner
David Wiley
Wallace Wilkens
Robert Williams
Ben James Winer
Eugene Winograd
Lawrence S. Wrightman
Robert S. Wyer
Leon J. Yarrow
Melvin Zax
Carl N. Zimet

VOLUME 79, 1973

Harold P. Van Cott
Managing Editor

Anita DeVivo
Executive Editor
and General Manager

Barbara Virgo Hobbs
Supervising Technical
Editor

Kathleen M. Springer
Technical Editor

Henry S. Salmon
Advertising and Business
Manager

Anne Redman
Subscription Manager

$12-50

PUBLISHED MONTHLY BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 SEVENTEENTH ST. N. W.
WASHINGTON, D. C. 20036

Copyright © 1973 by the American Psychological Association, Inc.

6.3.81

CONTENTS OF VOLUME 79

Anisfeld, Moshe, and Klenbort, Irene. On the functions of structural paraphrases: The view from the passive voice ... 117
Berkowitz, Leonard. Reactance and the unwillingness to help others ... 310
Berry, K. L. *See* Braucht, G. Nicholas
Birnbaum, Michael H. The devil rides again: Correlation as an index of fit ... 239
Blanchard, Edward B., and Young, Larry B. Self-control of cardiac functioning: A promise as yet unfulfilled ... 145
Bornstein, Harry, and Roy, Howard L. Comment on "Linguistic deficiency and thinking: Research with deaf subjects 1964–69" ... 211
Brainerd, Charles J. Judgments and explanations as criteria for the presence of cognitive structures ... 172
Brakarsh, Daniel. *See* Braucht, G. Nicholas
Braucht, G. Nicholas, Brakarsh, Daniel, Follingstad, Diane, and Berry, K. L. Deviant drug use in adolescence: A review of psychosocial correlates ... 92
Chapman, Jean P. *See* Chapman, Loren J. ... 380
Chapman, Loren J., and Chapman, Jean P. Problems in the measurement of cognitive deficit 380
Cole, Sherwood O. Hypothalamic feeding mechanisms and amphetamine anorexia ... 13
Cozby, Paul C. Self-disclosure: A literature review ... 73
Dallett, Janet. Theories of dream function ... 408
Darlington, Richard B. Comparing two groups by simple graphs ... 110
Depue, Richard A., and Fowles, Don C. Electrodermal activity as an index of arousal in schizophrenics ... 233
Dodd, David H., and Schultz, Roger F., Jr. Computational procedures for estimating magnitude of effect for some analysis of variance designs ... 391
Evans, Selby H. Reply to Sprott ... 180
Follingstad, Diane. *See* Braucht, G. Nicholas
Fowles, Don C. *See* Depue, Richard A.
Furth, Hans G. Further thoughts on thinking and language ... 215
Gibson, James J. Direct visual perception: A reply to Gyr ... 396
Gottlieb, Gilbert. Neglected developmental variables in the study of species identification in birds ... 362
Grings, William W. Cognitive factors in electrodermal conditioning ... 200
Hahn, William W. Attention and heart rate: A critical appraisal of the hypothesis of Lacey and Lacey ... 59
Healy, Alice F., and Jones, Cindy. Criterion shifts in recall ... 335
Hogan, Robert. Moral conduct and moral character: A psychological perspective ... 217
Jarrard, Leonard E. The hippocampus and motivation ... 1
Jones, Bill. Is there any proprioceptive feedback? Comments on Schmidt ... 386
Jones, Cindy. *See* Healy, Alice F.
Jones, Edward E., and Sigall, Harold. Where there is *ignis*, there may be fire ... 260
Jones, Stephen C. Self- and interpersonal evaluations: Esteem theories versus consistency theories ... 185
Klenbort, Irene. *See* Anisfeld, Moshe
Levine, David. *See* Miller, John K.
Lord, Frederic M. Testing if two measuring procedures measure the same dimension ... 71
Lorion, Raymond P. Socioeconomic status and traditional treatment approaches reconsidered 263
Lubow, Robert E. Latent inhibition ... 398
McWilliams, Spencer A., and Tuttle, Renee J. Long-term psychological effects of LSD ... 341
Miller, John K., and Levine, David. Correlation between genetically matched groups versus reliability theory: A reply to Jensen ... 142
Morgan, Wesley G. Nonnecessary conditions or useful procedures in desensitization: A reply to Wilkins ... 373
Morganstern, Kenneth P. Implosive therapy and flooding procedures: A critical review ... 318
Myers, Jerome L. *See* Perlmutter, Jane
Olton, David S. Shock-motivated avoidance and the analysis of behavior ... 243
Ostrom, Thomas M. The bogus pipeline: A new *ignis fatuus*? ... 252
Overall, John E., and Spiegel, Douglas K. Comment on Rawlings ... 164
Perlmutter, Jane, and Myers, Jerome L. Comparison of two procedures for testing multiple contrasts ... 181

Rajecki, D. W. Imprinting in precocial birds: Interpretation, evidence, and evaluation 48

Rawlings, Robert B., Jr. Comments on the Overall and Spiegel paper 168

Reid, Larry D. Processes of fear reduction in systematic desensitization: An addendum to Wilson and Davison (1971) 107

Roy, Howard L. *See* Bornstein, Harry

Schneider, David J. Implicit personality theory: A review: 294

Schultz, E. Fred, and Tapp, Jack T. Olfactory control of behavior in rodents 21

Schultz, Roger F., Jr. *See* Dodd, David H.

Scull, John W. The Amsel frustration effect: Interpretations and research 352

Shaffer, Juliet Popper. Defining and testing hypotheses in multidimensional contigency tables 127

Sigall, Harold. *See* Jones, Edward E.

Smith, I. Leon. Comment on Joe's "Comment on Overall and Spiegel's 'Least squares analysis of experimental data' " 170

Spiegel, Douglas K. *See* Overall, John E.

Sprott, D. A. Comment on misuse of covariance 180

Strauss, Milton, E. Behavioral differences between acute and chronic schizophrenics: Course of psychosis, effects of institutionalization, or sampling biases? 271

Swartz, Jon D. Gamble's review of the Holtzman Inkblot Technique: Corrections and clarifications 378

Tapp, Jack T. *See* Schultz, E. Fred

Treisman, Michel. Relation between signal detectability theory and the traditional procedures for measuring thresholds 45

Tuttle, Renee J. *See* McWilliams, Spencer A.

Wilkins, Wallace. Desensitization: A rejoinder to Morgan 376

Wood, Michael T. Power relationships and group decision making in organizations 280

Young, Larry B. *See* Blanchard, Edward B.

VOL. 79, NO. 1 JANUARY 1973

Psychological Bulletin

© 1973 by the American Psychological Association, Inc.

THE HIPPOCAMPUS AND MOTIVATION[1]

LEONARD E. JARRARD[2]

Carnegie-Mellon University

The available literature suggests that several different behavioral processes may be affected when the hippocampus is experimentally manipulated. In addition to convincing evidence that the structure is involved in human memory, and findings from research indicating a role in response perseveration, the evidence reviewed in this paper suggests that the hippocampus is involved in motivation. Specifically, it is proposed that the hippocampus functions as a negative feedback mechanism serving to inhibit the activity of lower structures more intimately involved in the control of nonspecific levels of activation and possibly some types of incentive motivation (thirst and hunger). Evidence is reviewed that supports the view that all regions of hippocampus are not functionally equal. This functional dissociation is characterized by (*a*) the anterodorsal CA1 cell fields that project via dorsal and postcommissural fornix to hypothalamus and thalamus being involved in incentive motivation, (*b*) CA3-CA4 cell fields that send axons through fimbria to the septal region being more concerned with the inhibiting of responses, and (*c*) all cell fields contributing to the influence that hippocampus has on nonspecific levels of activation.

Many investigators have thought of the functions of the hippocampus in terms of a unitary process. Several different points of view have been proposed: (*a*) The early formulation that the structure is concerned with the integration of emotional and motivational aspects of behavior (Kluver & Bucy, 1937; Papez, 1937). (*b*) The hippocampus plays a role in the mechanism of attention by inhibiting the orienting reflex (Grastyan, 1959). (*c*) The structure is primarily concerned with memory, more specifically, "recent" memory (Meissner, 1966; Milner, 1968; Milner & Penfield, 1955). (*d*) The hippocampus is involved in inhibiting learned and unlearned responses (Douglas, 1967; Kimble, 1968; Pribram, 1967). (*e*) Hippocampal activity is closely related to motor responses, especially voluntary types of movement (Vanderwolf, 1971). It has been difficult to obtain support for any one theory to the exclusion of the others since a variety of effects on behavior have been reported when this limbic structure is surgically ablated, or chemically and electrically stimulated. This is not surprising when one considers that damage to hippocampus has varied widely both in location and extent, and implanted electrodes often differ in location within hippocampus. Comparisons are further complicated by differences in behavioral testing procedures and differences in species that have been employed as subjects.

A careful review of the literature does not support the view that the hippocampus is concerned with a single function. Rather, available neuroanatomical, neurophysiological, and behavioral data suggest that several different behavioral processes may be affected when the structure is experimentally altered. There is convincing clinical evidence that the hippocampus is involved in human memory (Milner, 1968), and evidence from research suggests a role in response perseveration (see the reviews by Douglas [1967] and Kimble [1968]). It is the contention of the present

---

[1] This research was supported by National Science Foundation Grants GB-8044 and GB-30113.

[2] Requests for reprints should be sent to L. E. Jarrard, Department of Psychology, Washington and Lee University, Lexington, Virginia 24450.

No. 705(5)
4 | 1 | 74

writer that the hippocampus also has a significant influence on nonspecific changes in activation or arousal and possibly some specific types of incentive motivation (thirst and hunger). In addition, there is increasing evidence that different regions within hippocampus may be involved in some functions more than others. It is the purpose of the present paper to review the evidence supporting a motivational involvement of hippocampus and to speculate on the existence of functional dissociation within the structure.

Only the neuroanatomical, electrophysiological, and neuroendocrine details that are most relevant to present arguments are presented here since several excellent reviews are available (Green, 1964; Lorente de No, 1934; Stumpf, 1965). Most of the relevant behavioral data are from studies in which the rat has been used as a subject and the lesion technique employed to experimentally manipulate the hippocampus.

CONSIDERATIONS

Neuroanatomical

The major afferent and efferent connections of the hippocampus are with structures known to be important in motivation (Cragg, 1961; Green, 1964; Lorente de No, 1934; Nauta, 1958; Raisman, Cowan, & Powell, 1966). Because of the close anatomical relationship between hippocampus and hypothalamus, especially medial hypothalamus (Nauta, 1958), one might expect the hippocampus to exert an important influence on hypothalamic mechanisms controlling hunger, thirst, and spontaneous activity. Further, there are direct hippocampal efferents to the central grey midbrain substance (Nauta, 1958) and the septal region (Raisman et al., 1966).

Many of the early studies involved lesioning fornix with the assumption being that most, if not all, hippocampal efferents pass through fornix. However, there is neurophysiological evidence for major hippocampotemporal connections in cat and monkey (Green & Adey, 1956; Votaw, 1960). In addition, neuroanatomical evidence obtained by the present writer using Cebus monkey and the Nauta stain demonstrated the existence of extensive efferents from hippocampus to subiculum and entorhinal cortex (unpublished results). Hjorth-Simonsen (1971) has reported hippocampoentorhinal connections in the rat. These connections are apparently sparse since Raisman et al. (1966) did not find evidence for hippocampoentorhinal pathways using the Nauta stain, and Hjorth-Simonsen (1971) found it necessary to facilitate the search by employing preliminary operations on 8-day-old rats to deafferent hippocampus. Thus, the available evidence suggests that there may be important species differences in distribution of hippocampal efferents.

It has long been known that the hippocampus can be divided into several cytoarchitectonic fields (Lorente de No, 1933, 1934). In the excellent study by Raisman et al. (1966), it was shown with the Nauta stain that hippocampal cell fields in the rat have differential connections to extrahippocampal structures. They found that fibers in the dorsal fornix and the fimbria differ both in distribution and origin. The dorsal fornix was found to receive its fibers primarily from the anterior part of hippocampal field CA1 and to project via postcommissural fornix to the anterior thalamic nuclei and medial and lateral mammillary nuclei. Fields CA3 and CA4 distribute their fibers through the fimbria and precommissural fornix to the lateral septal nucleus and have reciprocal connections with the medial septal nucleus and nucleus of the diagonal band (Raisman, 1966). The posterior part of field CA1 has fibers distributed through both dorsal fornix and fimbria; thus, sending axons to both major areas that receive hippocampal efferents. On the basis of these data, Raisman et al. (1966) proposed a model in which there are two essentially distinct neuronal systems, one corresponding closely to the classical limbic system and including the anterior part of field CA1, and the other consisting of fields CA3 and CA4, medial and lateral septal nuclei, and nucleus of the diagonal band.

Electrophysiological

Several investigators have suggested that the hippocampus inhibits the reticular activating system (Adey, Segundo, & Livingston, 1957; Grastyan, 1959). Bremer, Stoupel, and VanReeth (1960) showed that hippocampal stimulation had an effect on cortical evoked potentials which was opposite to that of stimulation of the reticular formation, thus suggesting a close antagonistic interaction between the two brain areas. More recently, Redding (1967) demonstrated that the hippocampus has both an inhibitory influence on the brainstem reticular formation and a facilitatory effect on nonspecific thalamic systems.

The relations between hippocampal activity and various motivational states have been investigated. It was reported by Grastyan, Karmos, Vereczkey, and Kellenyi (1966) that pull (approach) and push (avoidance) reactions were regularly accompanied by hippocampal theta and desynchronization, respectively. Ford, Bremner, and Richie (1970) noted that electroencephalic activity in the hippocampal theta range was greater after 22 hours than after 3 hours of food deprivation. In other research hippocampal theta decreased during repeated exposure to a novel environment but was reinstated with both food and water deprivation and "blocked" during initiation of consummatory responses (Routtenberg, 1968). Routtenberg theorized that the hippocampus modulates two oscillating subcortical systems consisting of the reticular formation (activity in this system being signaled by hippocampal desynchronization) and the limbic system (activity accompanied by hippocampal theta).

Neuroendocrine

The hippocampus appears to have an important influence on the neuroendocrine system and thus homeostatic regulation of physiological processes. Neuroanatomical projections from hippocampus to the arcuate nucleus of the hypothalamus have been described (Nauta, 1956), and this cell group is involved in hormonal control. Mason (1958) studied the influence of various limbic structures on pituitary-adrenal activity and reported that destruction of hippocampus interrupted the usual diurnal rhythm of 17-hydroxycorticosteroids seen in normal monkeys. While urinary levels of this steroid normally fall during the night, the levels remained elevated in monkeys with hippocampal lesions. Other investigators (Endroczi & Lissak, 1960; Feldman, Wajshort, & Bernbaum, 1967; Knigge, 1961) looked at adrenalcorticosteroid and potassium output of the gastric secretory cells and concluded that the hippocampus has an inhibitory effect on secretion of these substances. In 1964 Green summarized the available evidence and stated that the hippocampus is inhibitory for the release of adrenocorticotropic hormone and possibly excitatory for the release of thyrotropic hormone (Green, 1964). In view of the increases in pituitary-adrenal cortical activity during conditions of arousal and alerting (Mason, 1968), and the normal inhibitory effect of hippocampus on this system, experimental manipulations of hippocampus should have an effect on behaviors that are influenced by endocrine system functioning.

BASIC MOTIVATIONAL PROCESSES

General Activity

One change in behavior frequently reported by investigators following hippocampal lesions in rats is an increase in general activity (Douglas & Isaacson, 1964; Jarrard & Bunnell, 1968; Kimble, 1963; Nadel, 1968; Sengstake, 1968; Teitelbaum & Milner, 1963). However, not all investigators have found significant increases (Boitano, Lubar, Auer, & Furnald, 1968; Douglas & Isaacson, 1964; Kaada, Rasmussen, & Kveim, 1961; Kim, 1960). These conflicting results are not surprising when one considers the multiple variations in activity-measuring devices and procedures that have been employed (see Strong & Jackson, 1970), and differences in size and location of the lesions. Another complicating factor is that most investigators have looked at activity for relatively short periods of time in novel situations. Animals with hippocampal lesions are often reported to habituate more slowly than controls (Jarrard & Korn, 1969; Leaton, 1965; Roberts, Dember, & Brodwick, 1962), and this has been interpreted as a deficit in response inhibition (Kimble, 1968). It is apparent that differences in general activity, habituation, and exploration would be confounded in these other studies.

Behavior in the home cage and novel open field has been studied by Jarrard (1968). Groups did not differ in the home cage during the day, but hippocampectomized rats were more active at night. (Similar results have been reported by Kim, Choi, Kim, Chang, Park, & Kang [1970].) In the open field, hippocampals were more active than controls but habituated at the same rate. In a replication and extension of this research, Jarrard and Simsak[3] found that the differential effect on day–night activity persisted 4 months after the operations.[4] Further, with continuous white light the usual 24-hour circadian rhythm increased the same as in control subjects (to approximately a 25½-hour rhythm), and hippocampals readjusted to the usual 12-hour-day–12-hour-night lighting regimen as quickly as controls. However, hippocampals were still more active during the time that control subjects were more active. Thus, it would appear that rats with lesions in hippocampus are more affected than controls by the novel, external stimuli found in the open field and by whatever the internal stimuli are that cause rats to be more active at night than during the day.

Several investigators have recorded activity of rats with hippocampal lesions under different deprivation conditions (Gotsick, 1969; Niki, 1962; Sengstake, 1968). Although hippocampals were usually more active than control subjects, the lack of differential effects with increasing deprivation agrees with the findings reported for normal rats (Bolles, 1967). Related is a study by Schmaltz and Isaacson (1967) in which hippocampals were found to be more active than controls before being tested in a bar-pressing task but not after eating the daily food ration.

When activity is recorded in situations that evoke goal-directed behavior, rats with hippocampal lesions are frequently more active than control subjects. Thus, increased intertrial activity during development of avoidance in the two-way shuttle box was noted by Green, Beatty, and Schwartzbaum (1967) for subjects with large bilateral hippocampal lesions but not subjects with less extensive damage. Kaplan (1968) reported no differences in acquisition of a conditioned freezing response, but during extinction subjects with hippocampal lesions were less perseverative than normal subjects, that is, they were more active.

Grant and Jarrard (1968) employed chemical stimulation of different hippocampal cell fields in rats and looked at activity, eating, and drinking. Stimulation with carbachol at all locations within hippocampus resulted in pronounced increases in activity with occasional motor seizures. Similar results have been reported by MacLean (1957) with electrical stimulation. Hyperactivity and sniffing resulted from low voltage stimulation, while motor seizures and catatonic responses occurred during hippocampal afterdischarges following higher voltage stimulation. Grastyan (1959) did not report hyperactivity in cats with low-level stimulation, but noted that stimulation was followed by rebound or release effects that were sufficient to activate previously inhibited conditioned responses. Since the hippocampus is particularly susceptible to epileptiform activity (Green, 1964), the increases in general activity during and following stimulation may have been the result of seizural discharges.

Reactivity to Stimuli

While it is not possible to know the extent to which the increases in activity described above may be the result of increases in reactivity, there have been several studies designed to look specifically at how animals with hippocampal lesions react to external stimuli. Hippocampal lesions do not result in increases in reactivity to presentation of either photic (Green et al., 1967; Kaplan, 1968) or auditory (Gotsick, 1969) stimuli unless the stimulation is intense (see Ireland & Isaacson, 1968). Kamback (1967) found slower habituation in hippocampals bar pressing for stimulus change under both ad-lib and food deprivation conditions. Because of the similarity in behavior of satiated hippocampals and food-deprived control subjects, a paralle was drawn between the effects of hippocampal lesions and the effects of food deprivation on habituation to novel stimulation.

---

[3] L. E. Jarrard & M. A. Simsak. Changes in activity, eating, and drinking following lesions of the hippocampus. Paper presented at the meeting of the Psychonomic Society, Saint Louis, October 1968.

[4] The persisting hyperactivity in the home cage probably accounts for the lighter body weight and higher levels of liver glycogen recently reported to exist in rats with hippocampal damage 8 months following the operation. H. M. Murphy, C. Wideman, & T. S. Brown. Liver glycogen levels in rats with limbic lesions. Paper presented at the meeting of the Psychonomic Society, San Antonio, October 1970.

In several studies reactivity to stimuli has been studied when the subjects are engaged in other ongoing behaviors. Thus, novel stimuli have been presented during drinking (Hendrickson, Kimble, & Kimble, 1969), barpressing (Kaplan, 1968), and performance in a straightway (Raphelson, Isaacson, & Douglas, 1965; Riddell, Rothblat, & Wilson, 1969; Wickelgren & Isaacson, 1963). Crowne and Riddell (1969) measured the cardiac component of orienting and found that hippocampal rats failed to show a normal response to novel stimuli when they were attending to other stimuli. The results of these studies show that subjects with hippocampal lesions are less distracted by extraneous stimulation when the subjects are engaged in goal-directed activity.

Several investigators have been concerned with the possibility that subjects are more affected by noxious stimuli following lesions of hippocampus. This is especially important since avoidance from shock has been used in many experiments. Although it has been reported that threshold for reaction to shock does not usually change, hippocampals overreact to shock intensities that are well above threshold (Evans & Jarrard, unpublished results; Schlosberg, Johnson, & Lash [5]). Blanchard and Fial (1968) determined shock thresholds for jump, flinch, vocalization, and crouching in rats with hippocampal lesions. While there were slight but significant differences in thresholds (hippocampals jumped at a lower shock level and vocalized at higher levels), the most important finding was decreased crouching following shock. With higher levels of shock, heart rate has been observed to increase in hippocampals following a single shock while the same conditions cause a heart rate decrease in control animals (Jarrard & Korn, 1969). McNew and Thompson (1966) observed that hippocampal subjects reacted more vigorously to shock during acquisition of an avoidance response. Thus, from the available evidence it would appear that differences in shock threshold are small but subjects with hippocampal lesions react more than controls to shock levels that are above threshold.

Drinking and Eating Behaviors

The hippocampus seems to have an influence on drinking and eating. Significant increases in ad-lib drinking have been reported following lesions of hippocampus (Glickman, Higgins, & Isaacson, 1970; Kimble & Coover, 1966); however, other investigators found no change in water consumption (Boitano et al., 1968). No differences have been found either in amount of water consumed or rate of drinking following 48 hours of deprivation (Gotsick, 1969; Kaada et al., 1961). Further, Milgram (1969b) was unable to elicit drinking of water following electrical stimulation of dorsolateral hippocampus, but he did find that the rats would drink a liquid diet and a saccharin solution. Perhaps the most consistent finding relating to water consumption is induction of drinking following chemical stimulation of hippocampus (MacPhail, 1968; Mountford, 1969). Fisher and Coury (1962) elicited drinking in rats with carbachol stimulation of dorsomedial hippocampus. More recently, Grant and Jarrard (1968) obtained drinking with carbachol stimulation of anterior CA1 hippocampal cell fields but not CA3-CA4 cells.

Rats with hippocampal lesions do not usually consume more food than control animals (Boitano et al., 1968; Gotsick, 1969; Jarrard, 1965; Niki, 1962), but significant increases have been reported (Kimble & Coover, 1966; Wishart, Brohman, & Mogenson, 1969). While there were no differences in total amount of food consumed in the study by Jarrard and Simsak (see Footnote 3), the pattern of eating at night changed from eating in "meals" (where a meal is defined as continuous eating for at least 2 minutes) to significantly shorter periods of eating interrupted by periods of activity.

Milgram (1969a) obtained changes in feeding behavior following electrical stimulation of dorsolateral hippocampus. Stimulated rats would eat, perform a food-motivated response, and learn to press a bar for food. It is interesting that evidence for facilitated feeding was not obtained during but only following termination of the stimulation. Thus, it was suggested that the hippocampus may normally inhibit feeding, and the facilitation was a result of rebound excitation of another structure. Attempts to switch the rebound feeding to drinking of water, gnawing, and mouse killing were not successful (Milgram, 1969b).

---

[5] A. J. Schlosberg, C. N. Johnson, & L. Lash. Response to shock by hippocampectomized rats. Paper presented at the meeting of the Eastern Psychological Association, Philadelphia, April 1969.

Increases in eating have been reported following chemical stimulation of hippocampus. Coury (1967) obtained eating in rats with norepinephrine stimulation of dorsomedial hippocampus but not anterior and lateral hippocampal areas. More recently, Grant and Jarrard (1968) found that stimulation with either norepinephrine or carbachol at all hippocampal cell fields increased food consumption over sham stimulation levels. The explanation for these discrepant results is not apparent.

Choice, Preference, and Competition

There do not appear to be any significant changes in simple and preferential consummatory behavior following damage to hippocampus. Beatty and Schwartzbaum (1968) and the present writer (unpublished results) found that rats with hippocampal lesions failed to exaggerate consummatory reactivity to positive (sucrose and saccharin) and negative (quinine) properties of solutions. Although changes in consummatory behavior were reported by Murphy and Brown (1970) following lesions of hippocampus, the authors considered the changes to be secondary to observed increases in activity.

Jarrard[6] employed a paired-comparison procedure that forced subjects to choose between being in a compartment with food, with another rat, or in a novel situation. Food-deprived rats with hippocampal damage spent more time with the food and were more active than control subjects.

MOTIVATION AND LEARNING

A problem frequently overlooked in experimentally testing the hypothesis that the hippocampus is primarily concerned with response inhibition or learning and memory is that differences in motivation (activity, hunger, thirst, reactions to novel and noxious stimuli) can result in differences in learning or performance of a learned task. One might expect facilitated or impaired performance to the extent that hippocampectomy results in animals emitting responses that are compatible or incompatible with learning the task.

Differences in drive do not usually affect acquisition of simple tasks in normal animals, but complex learning is affected (Berlyne, 1968; Bindra, 1959). In rats with damage to hippocampus there is usually no impairment in acquisition of simple tasks, including straight runway (Jarrard, Isaacson, & Wickelgren, 1964; Niki, 1962), bar pressing on a continuous reinforcement schedule for food or water (Jarrard, 1965; Schmaltz & Isaacson, 1967), and simple visual and spatial discriminations (Kimble, 1963; Kimble & Kimble, 1965). Musty (1969) did not find differences in a simple avoidance task, but hippocampals were impaired in acquisition of a more complex pattern discrimination. Deficits in hippocampectomized rats are usually found in acquisition of complex tasks such as mazes (Jarrard & Lewis, 1967; Kaada et al., 1961; Thomas, 1971), and monkeys with hippocampal lesions have trouble learning sequential responses (Kimble & Pribram, 1963) and delayed alternation (Rosvold & Szwarcbart, 1964).

Subjects with hippocampal lesions have been reported to press a lever at higher rates than control rats with "moderate" levels of deprivation but not at higher or lower levels (Musty, 1969). Kimble and Greene (1968) found little evidence for latent learning in hippocampectomized subjects. One would expect these results with highly motivated subjects since high drive has the effect of narrowing the range of cues to which subjects respond (Bruner, Matter, Papanek, 1955; Easterbrook, 1959; Telegdy & Cohen, 1971), and amount of latent learning is inversely related to strength of drive (Johnson, 1953). Reports of less distractibility in hippocampal subjects (Raphelson, Isaacson, & Douglas, 1965; Wickelgren & Isaacson, 1963) could be due to a narrowing of the range of stimuli to which highly motivated subjects attend.

---

[6] L. E. Jarrard. Strength of preference for food in rats with rhinencephalic lesions. Paper presented at the meeting of The Southern Society for Philosophy and Psychology, Atlanta, March 1965.

Since hippocampals react more than controls to above-threshold levels of shock (see above), one might expect differences in learning and performance when shock avoidance is used as the incentive. Blanchard and Fial (1968) suggested that an observed decreased crouching following shock may account for the often-reported passive avoidance deficit characteristic of rats with hippocampal lesions. Musty (1969) found an impairment of hippocampal subjects in pattern discrimination with high shock levels but not when low shock was employed. In a related unpublished study the present investigator found only a slight impairment in hippocampectomized subjects in reversal of a Y-maze position habit for food but a severe impairment when shock avoidance was used.

Several investigators have studied the involvement of hippocampus in learning by employing stimulation techniques. Correll (1957) used subthreshold electrical stimulation of hippocampus in cats while Grossman and Mountford (1964) induced spreading depression in hippocampus of rats during learning. The results of both studies were interpreted as suggesting that interference with hippocampal functioning may affect performance variables such as motivation rather than associative mechanisms related to acquisition or retention.

It is apparent from the above studies that a motivational explanation must be ruled out before impaired (or facilitated) performance of animals with hippocampal lesions can be attributed to deficits in other mechanisms.

FUNCTIONAL DISSOCIATION WITHIN HIPPOCAMPUS

One of the most important questions now facing investigators studying hippocampus is whether the complex structure is homogeneous in function with all parts being equally important in influencing behavior, or whether there are functional differences associated with known neuroanatomical differences. It has been assumed by many, especially those working primarily at a behavioral level, that the hippocampus functions as a unit. Thus, size and location of a lesion (or stimulating electrode) within hippocampus are not usually seen as being important. Many of the contradictions found in the current literature would probably be resolved if it could be shown that functional dissociation does exist.

Neuroanatomical (Nauta, 1956; Raisman et al., 1966), neurophysiological (Adey, 1961; Brazier, 1970; Elul, 1964; MacLean, 1959; Votaw & Lauer, 1963), and neurochemical (Purpura, 1959) evidence suggests that the hippocampus has a topical organization. As previously described, Raisman et al. (1966) proposed a model in which there are two neuronal systems—one primarily involving anterior CA1 cell fields and the second fields CA3 and CA4. The main problem in attempting to functionally test the model is the complex manner in which the various hippocampal fields are arranged. However, the rat is a good subject to use in attempts to lesion or stimulate selected fields since CA1 is an outermost layer, and the majority of fields CA3 and CA4 are located in the caudal half of the hippocampus (see Raisman, 1966). Although lesions to anterodorsal or posteroventral regions would not be limited exclusively to sites of origin of the two major projections, with these selective lesions there should be differences in the extent to which the two main projections are involved.

Several experiments have been reported in which lesions were limited to different parts of hippocampus (Jackson, 1968). In one of the earliest, Kimura (1958) damaged either the anterior or posterior hippocampus of rats and reported that posterior lesions interfered with acquisition of a passive avoidance response while anterior lesions had no effect. Rabe and Haddad (1969) and Kaada et al. (1962) found no evidence for differential effects of dorsal and ventral lesions on acquisition of two-way and passive avoidance responses, respectively. Nadel (1968) reported that ventral lesions affected rate of habituation of exploratory responses, retarded extinction of a conditioned emotional response, and facilitated acquisition of a one-way active avoidance response. Dorsal hippocampal lesions did not affect rates of habituation, but did facilitate acquisition of a conditioned emotional response, and retarded acquisition of a one-way active avoidance response. Further, rats with either dorsal or ventral hippocampal lesions were hyperactive in a novel situation.

There have been several attempts to study hippocampal functions by selectively damaging pathways associated with the structure. Liss (1968) limited lesions to either lateral fornix or posteroventral portions of hippocampus and found impaired one-way active and passive avoidance, and facilitated shuttlebox learning. Since the midline fornix remained intact, the author felt that disruption was primarily in fibers from posteroventral hippocampus and not the anterodorsal region. MacDougall, Van Hoesen, and Mitchell (1969) compared large fornix lesions that disrupted both dorsal fornix and fimbria fibers with smaller lesions restricted to the medial aspects of fornix. Marked deficits in performance of an operant response (differential reinforcement for low rates of responding) were found in subjects with complete lesions of fornix but not in subjects with limited dorsal fornix lesions. The authors suggested that the septal connections with hippocampal fields CA3-CA4 via fimbria are part of a system important in response inhibition and timing behavior. More recently, Thomas (1971) compared subtotal hippocampal lesions with fornix lesions and found that both groups of subjects were significantly impaired in retention and savings in relearning a complex maze. Damage to fornix, while not limited to postcommissural fornix, was more extensive in the descending columns.

It is possible that more precise experimental manipulations of the different hippocampal cell fields can be obtained with chemical and electrical stimulation. Grant and Jarrard (1968) studied the effects of adrenergic and cholinergic stimulation of anterodorsal and posteroventral CA1 and CA3-CA4 cell fields on drinking, eating, and activity. Control of drinking was found to be dissociated both on a neuroanatomical (limited to anterodorsal CA1 cell field) and biochemical (limited to cholinergic stimulation) basis. However, neuroanatomical location was not important in the effects on eating and activity. Interestingly enough, subjects were more active following carbachol stimulation at all hippocampal placements, thus suggesting a general increase in level of activation with no dissociation. Milgram (1969a, 1969b) reported significant increases in feeding behavior with electrical stimulation of lateral dorsal hippocampus. The effect was less pronounced or absent with stimulation of other hippocampal areas. The reason for the apparent conflict that exists between the effects of lesions and of chemical and electrical stimulation of selected hippocampal cell fields is not clear.

Most of the available evidence on functional dissociation in hippocampus has been obtained in the rat. However, there are several studies suggesting that different hippocampal regions in the cat are not functionally equal. Andy, Peeler, and Foshee (1967) found that cats with complete lesions and those with lesions limited to dorsal hippocampus were impaired in retention of an avoidance response, whereas subjects with ventral lesions were similar to controls. Similar effects of fornix and entorhinal lesions on learning and retention were reported by Uretsky and McLeary (1969). Differential effects with electrical stimulation of dorsal and ventral regions have been reported on attack behavior (Siegel & Flynn, 1968), and changes in blood pressure have been reported in cats (Andy, Webster, Mukawa, & Bonn, 1961) and monkeys (Votaw & Lauer, 1963). Further, various hippocampal regions in the cat differ in terms of electrical activity (Adey, 1961; MacLean, 1959). It is difficult to compare the results of the above investigations with those in which cats and those in which rats have been used as subjects since comparable data on projections of the various hippocampal cell fields do not exist for the two species. Even more important is the possible existence of major hippocampal entorhinal projections in the cat (see above).

CONCLUSIONS

It would appear that the hippocampus has an important influence on several behavioral processes. In addition to being important in response perseveration in the rat (Douglas, 1967; Kimble, 1968), the evidence reviewed in this paper suggests that the hippocampus is intimately involved in motivation. Specifically, it is proposed here that the hippocampus functions as a negative feedback mechanism serving to inhibit the activity of lower structures with which it is reciprocally connected. Especially important is the influence that the hippocampus has on structures controlling nonspecific levels of activation or arousal. It is hypothesized that experimental manipulations of hippocampus (lesions, chemical and electrical stimulation) result in nonspecific changes in activation or general drive that affect the way the subject reacts to the environment. Thus, following damage to hippocampus, level of activation or drive should be increased, "general motor readiness" (Bindra, 1968) should be enhanced, and under conditions of low stimulation the subject should be more responsive to both external and internal stimuli. Performance in learning tasks (especially complex tasks) should be affected since high levels of activation or drive serve to narrow the range of cues used by subjects (Easterbrook, 1959; Telegdy & Cohen, 1971) and complex tasks usually require difficult sensory discriminations.

While there is evidence to suggest that the hippocampus may also be involved in the modulation of some specific types of drive or incentive motivation (especially drinking and eating), the exact nature of the involvement is not clear. Conflicting results have been obtained concerning the effects of lesions on water and food consumption, but there are changes in the pattern of drinking and eating following lesions of hippocampus, and consistent increases in consumption are reported with chemical stimulation.

The neuroanatomical, neurophysiological, and behavioral evidence reviewed above would appear to support the hypothesis that there is functional dissociation within hippocampus in the rat. Specifically, the anterodorsal CA1 cell field that projects via dorsal and postcommissural fornix to hypothalamus and anterior nucleus of the thalamus seems to be involved in incentive motivation. Thus, experimental manipulations of this area have been reported to affect drinking (Grant & Jarrard, 1968) and eating (Milgram, 1969a, 1969b). The CA3-CA4 hippocampal cell fields that send axons through the fimbria to the septal area and diagonal band nuclei are more concerned with the inhibiting of responses or response perseveration (MacDougall et al., 1969). The evidence further suggests that both areas may have some functions in common, for example, the involvement in general arousal (Grant & Jarrard, 1968; Nadel, 1968). Thus, it is possible that all cell fields contribute to the influence that the hippocampus has on the undifferentiated state of activation or arousal.

The above formulations provide a basis for predictions concerning the behavioral effects of experimentally manipulating the hippocampus. Confirmation or refutation of these predictions should contribute to the development of a more comprehensive theory of hippocampal function.

REFERENCES

Adey, W. R. Studies of hippocampal electrical activity during approach learning. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning.* Oxford: Blackwell, 1961.

Adey, W. R., Segundo, J. P., & Livingston, R. B. Corticofugal influences on intrinsic brainstem conduction in cat and monkey. *Journal of Neurophysiology,* 1957, **20**, 1–16.

Andy, O. J., Peeler, D. F., Jr., & Foshee, D. P. Avoidance and discrimination learning following hippocampal ablation in the cat. *Journal of Comparative and Physiological Psychology,* 1967, **64**, 516–519.

Andy, O. J., Webster, C. L., Mukawa, J., & Bonn, P. Electrophysiological comparisons of the dorsal and ventral hippocampus. In, *Physiologia de l'hippocampe.* Paris: Du Centre National De La Recherche Scientifique, 1961.

Beatty, W. W., & Schwartzbaum, J. S. Commonality and specificity of behavioral dysfunctions following septal and hippocampal lesions in rats. *Journal of Comparative and Physiological Psychology,* 1968, **66**, 60–68.

Berlyne, D. E. Drive and arousal. *Neurosciences Research Program Bulletin,* 1968, **6**, 23–27.

Bindra, D. *Motivation: A systematic reinterpretation.* New York: Ronald Press, 1959.

Bindra, D. Drive, incentive-motivation and reinforcement. *Neurosciences Research Program Bulletin,* 1968, **6**, 67–72.

Blanchard, R. J., & Fial, R. A. Effects of limbic lesions on passive avoidance and reactivity to shock. *Journal of Comparative and Physiological Psychology,* 1968, **66**, 606–612.

Boitano, J. J., Lubar, J. F., Auer, J., & Furnald, M. Effects of hippocampectomy on consummatory behavior and movement-inhibition in rats. *Physiology and Behavior,* 1968, **3**, 901–906.

Bolles, R. C. *Theory of motivation.* New York: Harper & Row, 1967.

Brazier, M. A. B. Regional activities within the human hippocampus and hippocampal gyrus. *Experimental Neurology,* 1970, **26**, 354–368.

Bremer, F., Stoupel, N., & VanReeth, P. C. Nouvelles recherches sur la facilitation et l'inhibition des potentiels evoques corticaux dans l'eveil reticulaire. *Archives Italiennes de Biologie,* 1960, **98**, 229–247.

Bruner, J. S., Matter, J., & Papanek, M. L. Breadth of learning as a function of drive level and mechanization. *Psychological Review,* 1955, **62**, 1–10.

Correll, R. E. The effect of bilateral hippocampal stimulation on the acquisition and extinction of an instrumental response. *Journal of Comparative and Physiological Psychology,* 1957, **50**, 624–629.

Coury, J. N. Neural correlates of food and water intake in the rat. *Science,* 1967, **156**, 1763–1765.

Cragg, B. G. Olfactory and other afferent connections of the hippocampus in the rabbit, rat and the cat. *Experimental Neurology,* 1961, **3**, 588–600.

Crowne, D. P., & Riddell, W. I. Hippocampal lesions and the cardiac component of the orienting response in the rat. *Journal of Comparative and Physiological Psychology,* 1969, **69**, 748–755.

Douglas, R. J. The hippocampus and behavior. *Psychological Bulletin,* 1967, **67**, 416–442.

Douglas, R. J., & Isaacson, R. L. Hippocampal lesions and activity. *Psychonomic Science,* 1964, **1**, 187–188.

Easterbrook, J. A. The effect of emotion on cue utilization and organization of behavior. *Psychological Bulletin,* 1959, **66**, 183–201.

Elul, R. Regional differences in the hippocampus of the cat. I. Specific discharge patterns of the dorsal and ventral hippocampus and their role in generalized seizures. II. Projections of the dorsal and ventral hippocampus. *Electroencephalography and Clinical Neurophysiology,* 1964, **16**, 470–488, 489–502.

Endroczi, E., & Lissak, K. The role of mesencephalon and archicortex in the activation and inhibition of the pituitary adrenocortical system. *Acta Physiological Hungaria,* 1960, **17**, 39–55.

Feldman, S., Wajshort, J., & Bernbaum, D. Effect of combined brain stimulation on gastric secretion acidity and potassium concentration in cats. *Brain Research,* 1967, **4**, 103–106.

Fisher, A. E., & Coury, J. N. Cholinergic tracing of a central neural circuit underlying the thirst drive. *Science,* 1962, **138**, 691–693.

Ford, J. G., Bremner, F. J., & Richie, W. R. The effect of hours of food deprivation on hippocampal theta rhythm. *Neuropsychologia,* 1970, **8**, 65–73.

Glickman, S. E., Higgins, T. J., & Isaacson, R. L. Some effects of hippocampal lesions on the behavior of Mongolian gerbils. *Physiology and Behavior,* 1970, **5**, 931–938.

Gotsick, J. E. Factors affecting spontaneous activity in rats with limbic system lesions. *Physiology and Behavior,* 1969, **4**, 587–593.

Grant, L. E., & Jarrard, L. E. Functional dissociation within hippocampus. *Brain Research,* 1968, **10**, 392–401.

Grastyan, E. The hippocampus and higher nervous activity. In, M. A. B. Brazier (Ed.), *The central nervous system and behavior.* New York: Josiah Macy Foundation, 1959.

Grastyan, E., Karmos, G., Vereczkey, L., & Kellenyi, E. E. The hippocampal electrical correlates of the homeostatic regulation of motivation. *Electroencephalography and Clinical Neurophysiology,* 1966, **21**, 34–53.

Green, J. D. The hippocampus. *Physiological Review,* 1964, **44**, 561–608.

Green, J. D., & Adey, W. R. Electrophysiological studies of hippocampal connections and excitability. *Electroencephalography and Clinical Neurophysiology,* 1956, **8**, 245–262.

Green, R. H., Beatty, W. W., & Schwartzbaum, J. S. Comparative effects of septo-hippocampal and caudate lesions on avoidance behavior in rats. *Journal of Comparative and Physiological Psychology,* 1967, **64**, 444–452.

Grossman, S. P., & Mountford, H. Learning and extinction during chemically induced disturbance of hippocampal functions. *American Journal of Physiology,* 1964, **207**, 1387–1393.

Hendrickson, C. W., Kimble, R. J., & Kimble, D. P. Hippocampal lesions and the orienting response. *Journal of Comparative and Physiological Psychology,* 1969, **67**, 220–227.

Hjorth-Simonsen, A. Hippocampal efferents to the ipsilateral entorhinal area: An experimental study in the rat. *Journal of Comparative Neurology,* 1971, **142**, 417–437.

Ireland, L. C., & Isaacson, R. L. Reactivity in the hippocampectomized gerbil. *Psychonomic Science,* 1968, **12**, 163–164.

Jackson, W. J. A comment on "The hippocampus and behavior." *Psychological Bulletin,* 1968, **69**, 20–22.

Jarrard, L. E. Hippocampal ablation and operant behavior in the rat. *Psychonomic Science,* 1965, **2**, 115–116.

Jarrard, L. E. Behavior of hippocampal lesioned rats in home cage and novel situations. *Physiology and Behavior,* 1968, **3**, 65–70.

Jarrard, L. E., & Bunnell, B. N. Open-field behavior of hippocampal lesioned rats and hamsters. *Journal of Comparative and Physiological Psychology,* 1968, **66**, 500–502.

Jarrard, L. E., Isaacson, R. L., & Wickelgren, W. O. Effects of hippocampal ablation and intertrial interval on runway acquisition and extinction. *Journal of Comparative and Physiological Psychology,* 1964, **57**, 442–444.

Jarrard, L. E., & Korn, J. H. Effects of hippocampal lesions on heart rate during habituation and passive avoidance. *Communications in Behavioral Biology,* 1969, **3**, 141–150.

Jarrard, L. E., & Lewis, T. C. Effects of hippocampal ablation and intertrial interval on acquisition and extinction in a complex maze. *American Journal of Psychology,* 1967, **80**, 66–72.

Johnson, E. E. The role of motivational strength in latent learning. *Journal of Comparative and Physiological Psychology,* 1953, **45**, 526–530.

Kaada, B. R., Rasmussen, E. W., & Kveim, O. Effects of hippocampal lesions on maze learning and retention in rats. *Experimental Neurology,* 1961, **3**, 333–355.

Kaada, B. R., Rasmussen, E. W., & Kveim, O. Impaired acquisition of passive avoidance behavior by subcallosal septal, hypothalamic, and insular lesions in rats. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 661–670.

Kamback, M. Effect of hippocampal lesions and food deprivation on response for stimulus change. *Journal of Comparative and Physiological Psychology,* 1967, **63**, 231–235.

Kaplan, J. Approach and inhibitory reactions in rats after bilateral hippocampal damage. *Journal of Comparative and Physiological Psychology,* 1968, **65**, 274–281.

Kim, C. Nest building, general activity, and salt preference of rats following hippocampal ablation. *Journal of Comparative and Physiological Psychology,* 1960, **53**, 11–16.

Kim, C., Choi, H., Kim, J. K., Chang, H. K., Park, R. S., & Kang, I. Y. General behavioral activity and its component patterns in hippocampectomized rats. *Brain Research,* 1970, **19**, 379–394.

Kimble, D. P. The effects of bilateral hippocampal lesions in rats. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 273–283.

Kimble, D. P. The hippocampus and internal inhibition *Psychological Bulletin,* 1968, **70**, 285–295.

Kimble, D. P., & Coover, G. D. Effects of hippocampal lesions on food and water consumption in rats. *Psychonomic Science,* 1966, **4**, 91–92.

Kimble, D. P., & Greene, E. G. Absence of latent learning in rats with hippocampal lesions. *Psychonomic Science,* 1968, **11**, 99–100.

Kimble, D. P., & Kimble, R. J. Hippocampectomy and response perseveration in the rat. *Journal of Comparative and Physiological Psychology,* 1965, **60**, 474–476.

Kimble, D. P., & Pribram, K. H. Hippocampectomy and behavior sequences. *Science,* 1963, **139**, 824–825.

Kimura, D. Effects of selective hippocampal damage on avoidance behavior in the rat. *Canadian Journal of Psychology,* 1958, **12**, 213–218.

Kluver, H., & Bucy, P. C. "Psychic blindness" and other symptoms following bilateral lobectomy in Rhesus monkeys. *American Journal of Physiology,* 1937, **119**, 352–353.

Knigge, K. M. Adrenocortical response to immobilization in rats with lesions in hippocampus and amygdala. *Federation Proceedings,* 1961, **20**, 185. (Abstract)

Leaton, R. N. Exploratory behavior in rats with hippocampal lesions. *Journal of Comparative and Physiological Psychology,* 1965, **59**, 325–330.

Liss, P. Avoidance and freezing behavior following damage to the hippocampus or fornix. *Journal of Comparative and Physiological Psychology,* 1968, **66**, 193–197.

Lorente de No, R. Studies on the structure of the cerebral cortex. I. The area entorhinalis. *Journal of Psychologie and Neurologie,* 1933, **45**, 381–438.

Lorente de No, R. Studies on the structure of the cerebral cortex. II. Continuation of the study of the ammonic system. *Journal of Psychologie and Neurologie,* 1934, **46**, 113–177.

MacDougall, J. M., Van Hoesen, G. W., & Mitchell, J. C. Anatomical organization of septal projections in maintenance of DRL behavior in rats. *Journal of Comparative and Physiological Psychology,* 1969, **68**, 568–575.

MacLean, P. D. Chemical and electrical stimulation of hippocampus in unrestrained animals. II. Behavioral findings. *Archives of Neurology and Psychiatry,* 1957, **78**, 128–142.

MacLean, P. D. The limbic system with respect to two basic life principles. In M. A. B. Brazier (Ed.), *The central nervous system and behavior.* New York: Josiah Macy Foundation, 1959.

MacPhail, E. M. Effects of intracranial cholinergic stimulation in rats on drinking, EEG, and heart rate. *Journal of Comparative and Physiological Psychology,* 1968, **65**, 42–49.

Mason, J. W. The central nervous system regulation of ACTH secretion. In H. H. Jasper (Ed.), *Reticular formation of the brain.* Boston: Little, Brown, 1958.

Mason, J. W. A review of psychoendocrine research on the pituitary adrenal cortical system. *Psychosomatic Medicine,* 1968, **30**, 576–607.

McNew, J. J., & Thompson, R. Role of the limbic system in active and passive avoidance conditioning in the rat. *Journal of Comparative and Physiological Psychology,* 1966, **61**, 173–180.

Meissner, W. M. Hippocampal functions in learning. *Journal of Psychiatric Research,* 1966, **4**, 235–304.

Milgram, N. W. Effect of stimulation on feeding in the rat. *Physiology and Behavior,* 1969, **4**, 665–670. (a)

Milgram, N. W. Motivational specificity of dorsolateral hippocampal stimulation. *Communications in Behavioral Biology,* 1969, **4**, 1–5. (b)

Milner, B. Disorders of memory after brain lesions in man. *Neuropsychologica,* 1968, **6**, 175–179.

Milner, B., & Penfield, W. The effect of hippocampal lesions on recent memory. *Transactions of the American Neurological Association,* 1955, **80**, 42–48.

Mountford, D. Drinking following carbachol stimulation of hippocampal formation or lateral ventricles. *Psychonomic Science,* 1969, **16**, 124–125.

Murphy, H. M., & Brown, T. S. Effects of hippocampal lesions on simple and preferential consummatory behavior in the rat. *Journal of Comparative and Physiological Psychology.* 1970, **72**, 404–415.

Musty, E. On hippocampal modulation of motivated behavior in the rat. *Dissertation Abstracts,* 1969, **30**, 414.

Nadel, L. Dorsal and ventral hippocampal lesions and behavior. *Physiology and Behavior,* 1968, **3**, 891–900.

Nauta, W. J. H. An experimental study of the fornix system in the rat. *Journal of Comparative Neurology,* 1956, **104**, 247–270.

Nauta, W. J. H. Hippocampal projections and related neural pathways to the mid-brain in the cat. *Brain,* 1958, **81**, 319–340.

Niki, H. The effects of hippocampal ablation on the behavior in the rat. *Japanese Psychological Research,* 1962, **4**, 139–153.

Papez, J. W. A proposed mechanism of emotion. *Archives of Neurology and Psychiatry,* 1937, **38**, 725–743.

Pribram, K. H. The limbic system, efferent control of neural inhibition and behavior. In W. R. Adey & T. Tokezane (Eds.), *Progress in brain research.* Vol. 27. *Structure and function of the limbic system.* New York: Elsevier Publishing, 1967.

Purpura, D. P. Discussion In M. A. B. Brazier (Ed.), *The central nervous system and behavior.* New York: Josiah Macy Foundation, 1959.

Rabe, A., & Haddad, R. K. Acquisition of two-way shuttle-box avoidance after selective hippocampal lesions. *Physiology and Behavior,* 1969, **4**, 319–323.

Raisman, G. The connections of the septum. *Brain,* 1966, **89**, 317–348.

Raisman, G., Cowan, W. M., & Powell, R. P. S. An experimental analysis of the efferent projection of the hippocampus. *Brain,* 1966, **89**, 83–108.

Raphelson, A. C., Isaacson, R. L., & Douglas, R. J. The effect of distracting stimuli on the runway performance of limbic damaged rats. *Psychonomic Science,* 1965, **3**, 483–484.

Redding, F. K. Modification of sensory cortical evoked potentials by hippocampal stimulation. *Electroencephalography and Clinical Neurophysiology,* 1967, **22**, 74–83.

Riddell, W. I., Rothblat, L. A., & Wilson, W. A., Jr. Auditory and visual distraction in hippocampectomized rats. *Journal of Comparative and Physiological Psychology,* 1969, **67**, 216–219.

Roberts, W., Dember, W. N., & Brodwick, M. Alternation and exploration in rats with hippocampal lesions. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 695–700.

Rosvold, H. E., & Szwarcbart, M. K. Neural structures involved in delayed-response performance. In J. M. Warren & K. Akert (Eds.), *The frontal granular cortex and behavior.* New York: McGraw-Hill, 1964.

Routtenberg, A. Hippocampal correlates of consummatory and observed behavior. *Physiology and Behavior,* 1968, **3**, 533–535.

Schmaltz, L. W., & Isaacson, R. L. Effect of bilateral hippocampal destruction on the acquisition and extinction of an operant response. *Physiology and Behavior,* 1967, **2**, 291–298.

Sengstake, C. B. Habituation and activity patterns of rats with large hippocampal lesions under various drive conditions. *Journal of Comparative and Physiological Psychology,* 1968, **65**, 504–506.

Siegel, A., & Flynn, J. P. Differential effects of electrical stimulation and lesions of the hippocampus and adjacent regions upon attack behavior in cats. *Brain Research,* 1968, **7**, 252–267.

Strong, P. N., & Jackson, W. J. Effects of hippocampal lesions in rats on three measures of activity. *Journal of Comparative and Physiological Psychology,* 1970, **70**, 60–65.

Stumpf, C. H. Drug action on the electrical activity of the hippocampus. *International Review of Neurology,* 1965, **8**, 77–138.

Teitelbaum, H., & Milner, P. Activity changes following partial hippocampal lesions in rats. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 284–289.

Telegdy, G. A., & Cohen, J. S. Cue utilization and drive level in albino rats. *Journal of Comparative and Physiological Psychology,* 1971, **75**, 248–253.

Thomas, G. J. Maze retention by rats with hippocampal lesions and with fornicotomies. *Journal of Comparative and Physiological Psychology,* 1971, **75**, 41–49.

Uretsky, E., & McCleary, R. A. Effect of hippocampal isolation on retention. *Journal of Comparative and Physiological Psychology,* 1969, **68**, 1–8.

Vanderwolf, C. H. Limbic-diencephalic mechanisms of voluntary movement. *Psychological Review,* 1971, **78**, 83–113.

Votaw, C. L. Certain functional and anatomical relations of the Cornu ammonis of the Macaque monkey. II. Anatomical relations. *Journal of Comparative Neurology,* 1960, **114**, 283–293.

Votaw, C. L., & Lauer, E. W. Blood pressure, pulse, and respiratory changes produced by stimulation of the hippocampus of the monkey. *Experimental Neurology,* 1963, **7**, 502–514.

Wickelgren, W. O., & Isaacson, R. L. Effect of the introduction of an irrelevant stimulus on runway performance of the hippocampectomized rat. *Nature,* 1963, **200**, 48–50.

Wishart, T., Brohman, L., & Mogenson, G. Effects of lesions of the hippocampus and septum on hoarding behavior. *Animal Behavior,* 1969, **17**, 781–784.

(Received June 24, 1971)

Psychological Bulletin
1973, Vol. 79, No. 1, 13–20

HYPOTHALAMIC FEEDING MECHANISMS AND AMPHETAMINE ANOREXIA[1]

SHERWOOD O. COLE[2]

Rutgers—The State University, Camden

The hypothalamic regulation of feeding is discussed in terms of the ventromedial hypothalamic "satiety center" and the lateral hypothalamic "feeding center." Two hypotheses concerning the anorexic properties of amphetamine are considered in terms of the drug's interaction with hypothalamic centers—a ventromedial-hypothalamic-mimicking hypothesis and a lateral-hypothalamic-blocking hypothesis. A consideration of these hypotheses suggests that the drug influences both centers, and that additional sites in the cerebral cortex and brainstem also may serve ancillary functions in mediating the drug's effect on feeding. It is suggested that future research should attempt to clarify the mediational processes underlying the drug's effect on feeding and nonfeeding performance and, in those instances where such processes overlap, to specify the nature of the overlap.

The general depressant effect of amphetamine on feeding and food-motivated behavior is well established (Cole, 1967, 1970). Carlton (1963) has suggested that amphetamine anorexia may be due to an increase in activation to the degree that responses incompatible with those required in feeding occur and thus result in a decrease in feeding. A second peripheral view of the action of amphetamine on feeding (Mogenson, 1968) suggests that the drug reduces liver and muscle glycogen and thereby increases the intracellular glucose in hepatic cells that causes gluco-receptors in the liver to signal a state of satiety. Such peripheral interpretations must be given proper consideration, but they do not explain the changing effectiveness of amphetamine on feeding following central nervous system lesions (Cole, 1967, 1968). Furthermore, amphetamine has been observed to depress feeding in the complete absence of peripheral changes in the circulatory system (Nathanson, 1937) and after interruption of nerve pathways regulating the gastro-intestinal tract (Harris, Ivy, & Searle, 1947).

The purpose of the present paper is (*a*) to review briefly the hypothalamic mechanisms regulating feeding and (*b*) to consider the anorexic action of amphetamine in terms of the drug's interaction with such mechanisms.

HYPOTHALAMIC FEEDING MECHANISMS[3]

Two specific regions of the hypothalamus are of major importance in the regulation of feeding—the ventromedial and lateral regions. These regions have been referred to as the "satiety center" and "feeding center" of the hypothalamus, respectively (Anand & Brobeck, 1951).

Ventromedial Hypothalamic Satiety Center[4]

Brobeck, Tepperman, and Long (1943) demonstrated that electrolytic lesions in the ventromedial hypothalamus produced an increase in feeding by rats. Several subsequent studies (Brooks, Lockwood, & Wiggins, 1946; Liebelt & Perry, 1957; Marshall, Barrett, &

---

[1] Preparation of this paper was supported, in part, by a Faculty Fellowship from the Rutgers University Research Council.

[2] The author thanks Dr. John R. Brobeck for stimulating discussions on the topic and for his reading of the original manuscript.

Reprint requests should be sent to Sherwood O. Cole, Department of Psychology, Rutgers University, Camden, New Jersey 08102.

[3] For an updated review on the hypothalamic control of feeding that appeared during the preparation of this paper, see Hoebel (1971).

[4] Reynolds (1963, 1965) has questioned the existence of a specific ventromedial hypothalamic "satiety center" and has hypothesized that electrolytic lesions in this area deposit metallic ions that produce scar tissue and have irritating properties to adjacent lateral hypothalamic areas. However, there is evidence that hyperphagia can be produced by destruction of the ventromedial hypothalamus in the absence of irritative effects (Gold, 1970a; Hoebel, 1965).

Mayer, 1955; Wagner & de Groot, 1963) supported these findings with bilateral lesions of the ventromedial hypothalamus, either electrolytically or chemically introduced, enormously exaggerating the feeding of experimental animals (hypothalamic hyperphagia). Such a lack of inhibition to eat with ventromedial hypothalamic lesions apparently involved a failure to adjust food intake to the energy demands of the body (Brobeck, 1946; Strominger, Brobeck, & Cort, 1953). Recent evidence suggests that although both male and female rats demonstrate an acute dynamic period immediately following surgery, followed by a more chronic dynamic period (Balagura & Devenport, 1970), the actual amount of hyperphagia resulting from ventromedial hypothalamic lesions may vary with sex (Balagura & Devenport, 1970; Cox, Kakolewski, & Valenstein, 1969). However, such differences may be due to a failure to equate sexes for weight (Gold, 1970b) or to a failure to take into consideration the influence of day–night differences in sexual adjustment (Balagura & Devenport, 1970). Although studies by Hetherington (1943, 1944) suggested that the critical region involved in the production of hyperphagia is the ventromedial nucleus, Graff and Stellar (1962) observed hyperphagia following bilateral hypothalamic lesions which spared the ventromedial nucleus but destroyed the area just posterior to it. Evidence for the satiety function of the ventromedial hypothalamus also was found by Wyrwicka and Dobrzecka (1960), with electrical stimulation of the ventromedial hypothalamus inhibiting the food intake of experimental subjects.

Lateral Hypothalamic Feeding Center

In a study by Delgado and Anand (1953), electrical stimulation of the nucleus of the lateral hypothalamus by means of chronically implanted electrodes produced an increase in food consumption; however, this increase was observed only after several days of daily 1-hour stimulation periods. Recent evidence (Devor, Wise, Milgram, & Hoebel, 1970) suggests that such behavior elicited by electrical stimulation is neither rigid nor undirected, thus supporting the view that lateral hypothalamic stimulation can trigger neural substrates that regulate feeding. Additional support for the facilitatory role of the lateral hypothalamus in regulating feeding was found in studies by Teitelbaum and Stellar (1954), by Anand and Brobeck (1951), and by Feldman, Larsson, Dimick, and Lepkovsky (1957) where bilateral lesions of the lateral hypothalamus resulted in either hypophagia or aphagia. However, there has been some question as to whether this deficit is motivational (Rodgers, Epstein, & Teitelbaum, 1965) or due to a disruption of motor systems involved in feeding (Morrison, 1968). On a few occasions, experimental subjects made hypophagic or aphagic by lateral hypothalamic lesions have been observed to recover eating spontaneously, but such subjects usually starve to death or recover only with special diets (Teitelbaum & Stellar, 1954; Williams & Teitelbaum, 1959). Teitelbaum and Epstein (1962) have reported a comprehensive study of the recovery stages of feeding following lateral hypothalamic lesions. Recent evidence also suggests that the postoperative recovery of lateral-hypothalamus-lesioned subjects depends upon preoperative feeding experience (Di Cara, 1970) and preoperative body weight (Powley & Keesey, 1970), and that there may be a sex difference in the recovery from such lesions (Cox & Kakolewski, 1970).

Additional Considerations

Anand and Brobeck (1951) demonstrated that bilateral lesions involving the ventromedial hypothalamic region or the region between it and the lateral region (but not involving the lateral region) produced hyperphagia and obesity. This hyperphagia ceased when the lateral hypothalamic regions were destroyed bilaterally. The authors interpreted the findings to indicate that the ventromedial hypothalamic region is capable of exerting its inhibitory influence over the lateral hypothalamic facilitatory activity through fibers that run laterally into the lateral region. Other evidence by Albert and Storlien (1969), by Gold (1970a), and by Arees and Mayer (1967) suggest that fibers do interact between the ventromedial and the lateral hypothalamus in the regulation of feeding. Evidence also suggests that the hypothalamic "centers" interact with other cortical regions in the regulation of feeding. Kirschbaum

(1951) observed some marked changes in the hunger and eating of patients with various types of organic lesions of the cerebral cortex. Also, Pribram and Bagshaw (1953) observed changes in feeding following ablation of the temporal polar-amygdaloid cortex, and Fuller, Rosvold, and Pribram (1957) observed striking increases in food intake following amygdalectomy. Recent evidence suggests that changes in feeding following damage to temporal lobe sites may be secondary in nature rather than due to incentive shifts, and that such sites have their influence on feeding by a modulation of the activity of cells in the hypothalamus (Murphy & Brown, 1970; Sclafani, Belluzzi, & Grossman, 1970; White & Fisher, 1969). Other evidence based on the similar effects of chemical stimulation of the amygdala and hypothalamus on feeding (Grossman, 1960, 1964) is consistent with the view that these sites interact in the regulation of feeding.

HYPOTHALAMIC MECHANISMS UNDERLYING ACTION OF AMPHETAMINE

Two basic hypotheses concerning the mechanisms underlying the depressant effects of amphetamine on feeding have been suggested by experimental findings—a ventromedial-hypothalamus-mimicking hypothesis and a lateral-hypothalamus-blocking hypothesis. An attempt is made to evaluate the merits and limitations of these hypotheses in handling the experimental data.

Ventromedial Hypothalamus-Mimicking Hypothesis

In a study by Brobeck, Larsson, and Reyes (1956), electrical recordings were taken in different subcortical regions of the brain in response to amphetamine injected intravenously in anesthetized cats. Since the ventromedial hypothalamic region responded to the injections with increased electrical activity, the results suggested that amphetamine depressed eating by chemically mimicking the activity of the ventromedial hypothalamic "satiety center." More recently, Krebs, Bindra, and Campbell (1969) found that amphetamine increases the firing rate of single neural units in the ventromedial hypothalamus, and Carr and Moore (1969) demonstrated that amphetamine releases norepinephrine from the ventromedial hypothalamus. Since Foxwell, Funderburk, and Ward (1969) found that another anorexic agent (fenfluramine) increases electrical activity in the ventromedial hypothalamus, and Wade and Zucker (1970) demonstrated that an estrogen stimulation of the ventromedial hypothalamus also decreases feeding, an application of the ventromedial-hypothalamus-mimicking hypothesis apparently is not amphetamine specific.

One basic assumption of the ventromedial-hypothalamus-mimicking hypothesis is that if the ventromedial hypothalamic region is destroyed by lesioning, the depressant effects of amphetamine on feeding should be attenuated or eliminated. A study by Sharp, Neilson, and Porter (1962), in which ventromedial hypothalamic lesions attenuated the depressant action of amphetamine on feeding in cats, is consistent with this assumption. However, other studies (Epstein, 1959; Reynolds, 1959) have found quite the opposite effect; that is, the lesions exaggerated the depressant effect of the drug on feeding.[5] Although these latter findings suggest the need for rejecting the ventromedial-hypothalamus-mimicking hypothesis, a consideration of additional facts might still allow one to accept such a view with some qualifications. Margules (1969, 1970) has suggested that the exaggerated amphetamine effect following ventromedial hypothalamic lesions in rats may represent a denervation supersensitivity of remaining alpha-adrenergic sites. Thus, amphetamine may depress feeding by mimicking the activity of the ventromedial hypothalamic "satiety center" in the intact brain, but the assumption underlying the reversal of this action with ventromedial hypothalamic lesions (i.e., attenuation or elimination of the drug's effect) does not necessarily hold because of a changed sensitivity of remaining sites. Provided one accepts the view that amphetamine does not mediate its effect on feeding solely by mimicking the activity of the ventromedial hypothalamic "satiety center," the concept of denervation hypersensitivity has some merit in explaining the exaggerated action of the drug following ventromedial hypothalamic lesions. One might also suggest that other findings of an exaggerated effect of amphetamine on feeding following anterior hypothalamic lesions (Cole, 1966) are due to a similar denervation hypersensitivity. The concept of denervation hypersensitivity does not explain the findings of Sharp et al. (1962) that ventromedial hypothalamic lesions attenuated the effects of amphetamine on feeding. Since cats were used in the study by Sharp et al., in which support was found for the ventromedial-hypothalamus-mimicking hypothesis, and rats were used in the studies that did not support the hypothesis (Epstein, 1959; Reynolds, 1959), a species-specific explanation of the discrepant findings cannot be ruled out. However, a species-specific explanation does not hold when considering the fact that in the intact brain the drug has been observed to increase the firing rate of neural units in the ventromedial hypothalamus of cats (Brobeck et al., 1956) and of rats (Krebs et al., 1969) in much the same manner. As an alternative explanation of such discrepant findings, Carlisle (1964) has suggested that the study by Sharp et al. may not be a valid test of the changing effectiveness of amphetamine following ventromedial hypothalamic lesions, since food consumption was measured over a 24-hour period rather than during a restricted period representing the peak action of the drug.

---

[5] Stowe and Miller (1957) previously had found a similar exaggerated effect of amphetamine following ventromedial hypothalamic lesions. The discrepancy in these findings pertaining to the effects of amphetamine following ventromedial hypothalamic lesions has been pointed out by Meyer and Meyer (1963).

Lateral Hypothalamus-Blocking Hypothesis

Booth (1968) has demonstrated that cannula-guided chemical stimulation of the lateral hypothalamus resulted in a consistency of correlation between the effectiveness of noradrenalin to elicit eating and the effectiveness of amphetamine to depress eating. These results suggested that amphetamine mediates its anorexic action on feeding by the blocking of adrenergic activity in the lateral hypothalamic "feeding center." Basic to this blocking may be a direct stimulation of beta-adrenergic cells in the lateral hypothalamus (Leibowitz, 1970) or an increase in the threshold of excitability in the "feeding center," since the drug has been found to increase the threshold for eating by stimulation of the lateral hypothalamus (Miller, 1960; Stark & Totty, 1967). Additional evidence that the anorexic effects of amphetamine are counteracted by electrical stimulation of the lateral hypothalamus (Thode & Carlisle, 1968) and that amphetamine decreases the amplitude of elicited lateral hypothalamic activity (Reiter, 1970) also are consistent with the view that the drug depresses feeding by decreasing lateral hypothalamic excitability.

As was true with the ventromedial-hypothalamus-mimicking hypothesis, a basic assumption underlying the lateral-hypothalamus-blocking hypothesis is that if the lateral hypothalamic region is destroyed by lesioning, the depressant effects of amphetamine on feeding should be attenuated or eliminated. The findings by Carlisle (1964) that the depressant effect of amphetamine was attenuated in lateral-hypothalamus-lesioned subjects who had recovered from postoperative aphagia-adipsia support such an assumption. Since the lateral-hypothalamus-lesioned subjects and the control subjects did not differ significantly with respect to the action of the drug on water intake, the changing effectiveness of amphetamine following such lesions cannot be attributed to indirect influences on feeding behavior.

In addition, the lateral-hypothalamus-blocking hypothesis would seem to provide a meaningful organization of the exaggerated effects of amphetamine on feeding following ventromedial hypothalamic lesions (Epstein, 1959; Reynolds, 1959) in terms of changing threshold properties of the lateral hypothalamic "feeding center." There is evidence that electrolytic ventromedial hypothalamic lesions produce the typically observed hyperphagic syndrome by the releasing of inhibitory influences through fibers running laterally into the lateral hypothalamus from the ventromedial hypothalamus (Anand & Brobeck, 1951). Thus, the exaggerated effect of amphetamine may be due to changes in the reactive sensitivity of the lateral hypothalamic "feeding center" produced by the release of ventromedial hypothalamic restraint, which, in turn, increases the effectiveness of the drug in raising the lateral hypothalamic threshold of excitability. Although a similar but less direct "release of restraint" may underly the exaggerated action of the drug following anterior hypothalamic lesions (Cole, 1966), the specific nature of the release still remains to be determined. Since anterior-hypothalamus-lesioned subjects do not characteristically overeat unless they are heat stressed (Hamilton, 1963), the exaggerated action of the drug cannot be attributed to a simple "release of restraint" on feeding motivation as might well be done with changes produced by ventromedial hypothalamic lesions. However, fibers of the medial forebrain bundle are located in the anterior hypothalamus and serve important functions in regulating the activity of the lateral hypothalamic "feeding center" (Morgane, 1961a, 1961b); thus, a "release of restraint" on such connections with the "feeding center" resulting from anterior hypothalamic lesions might produce indirectly the same kind of exaggerated effect on amphetamine's action as is produced more directly by ventromedial hypothalamic lesions. The lateral-hypothalamus-blocking hypothesis would seem to provide a meaningful organization of the exaggerated effect of amphetamine following ventromedial hypothalamic lesions for another reason as well. The proposed site of amphetamine's action (lateral hypothalamic "feeding center") is intact and could become hypersensitive or less restrained, thereby mediating an increased effectiveness of the drug. In contrast, an organization of such evidence by the ventromedial-hypothalamus-mimicking hypothesis is less clear, since the very structure hypothesized to mediate the drug's action (the ventromedial hypothalamus) is no longer intact.

Further Comments on the Mediation of Amphetamine Anorexia

Although it would be convenient and certainly parsimonious to find that amphetamine mediates its depressant effect on feeding by either mimicking the ventromedial hypothalamic "satiety center" or by blocking activity in the lateral hypothalamic "feeding center," the evidence suggests that both regions are involved in the mediation of the drug's anorexic properties. While the evidence obtained from the intact brain (via recording and/or stimulation) supports both of the previous hypotheses, the lesioning evidence less clearly supports either. The findings of Sharp et al. (1962) support the ventromedial-hypothalamus-mimicking hypothesis, and the findings of Carlisle (1964) support the lateral-hypothalamus-blocking hypothesis; however, the exaggerated action of the drug following ventromedial hypothalamic lesions (Epstein, 1959; Reynolds, 1959) might possibly be explained by either hypothesis, provided one is willing to entertain some additional assumptions. Furthermore, as pointed out by Kreb et al. (1969), even the evidence that favors the ventromedial-hypothalamus-mimicking hypothesis does not necessarily rule out the possibility that the drug produces its depressant effect on feeding by acting on the lateral hypothalamic "feeding center."

Although studies (Glick, 1969; Glick & Jarvik, 1970; Jonason, Lauber , Robbins, Meyer, & Meyer, 1970; Novick & Pihl, 1969) have investigated the combined effects of amphetamine and lesions in sites other than the hypothalamus on learning and performance, little evidence is available on the role such sites play in mediating the anorexic action of the drug. What evidence there is suggests that the role of cortical and other subcortical sites in regulating the drug's effect on feeding parallels the role of such sites in the regulation of general feeding. Harris et al. (1947) observed that human subjects with frontal lobotomies were less susceptible to the anorexic properties of amphetamine than were normal subjects, and Andersson and Larsson (1957) observed similar changes in the depressant effect of the drug on feeding in dogs following prefrontal lobotomies. Recently, Carey and Salim (1970) found that septal lesions slightly attenuated the anorexic effect of amphetamine, but the attenuation did not reach statistical significance. Since Carlisle and Reynolds (1961) observed an exaggerated effect of amphetamine on feeding following brainstem (area postrema) lesions, sites downstream to the hypothalamus may also serve important ancillary functions in the total mediation of the drug's action. Ruch, Patton, and Brobeck (1942) have identified fibers that descend from the lateral hypothalamic "feeding center" and exert their influence upon internuncial motor neurons and nerves involved in the execution of feeding: Thus, it is possible that the exaggerated effect of the drug in this latter case is due to the "release of restraint" on executive functions of feeding.

Concluding Comments

Amphetamine is similar in structure to the catecholamines and mimics many of their central effects (Carlton, 1963). Furthermore, the importance of such adrenergic systems to the effects of the drug on nonfeeding performance is well established (Anden, Carlsson, & Haggendal, 1969; Maickel, Cox, Miller, Segal, & Russell, 1969; Weiss & Laties, 1969).

Although in some instances it may be possible to separate the brain mechanisms underlying the effects of amphetamine on feeding from those underlying the effects of the drug on nonfeeding tasks, this would not always seem to be possible. Separate mechanisms have been suggested in the case of general motor behavior and feeding (Cole, 1966; Van Rossum & Simons, 1969); however, in the case of operant performance, the brain sites regulating self-stimulation and feeding may overlap considerably (Gluckman & Baum, 1969; Hoebel & Teitelbaum, 1962; Khavari, 1969; Olds, 1962; Pickens & Harris, 1968; Poschel & Ninteman, 1966; Stein, 1964).

Future research should attempt to clarify further the role of hypothalamic mechanisms in mediating the anorexic properties of amphetamine and to identify more specifically the adrenergic systems underlying the drug's action on nonfeeding performance. In those instances where there is an overlap of such mediational processes, attempts should be made to identify the structural and/or functional nature of the overlap.

REFERENCES

Albert, D. J., & Storlien, L. H. Hyperphagia in rats with cuts between the ventromedial and lateral hypothalamus. *Science*, 1969, **165**, 599–600.

Anand, B. K., & Brobeck, J. R. Hypothalamic control of food intake in rats and cats. *Yale Journal of Biology and Medicine*, 1951, **24**, 123–138.

Anden, N. E., Carlsson, A., & Haggendal, J. Adrenergic mechanisms. *Annual Review of Pharmacology*, 1969, **9**, 119–134.

Andersson, B., & Larsson, S. Water and food intake and the inhibitory effect of amphetamine on drinking and eating before and after pre-frontal lobotomy in dogs. *Acta Physiologica Scandinavica*, 1957, **38**, 22–30.

Arees, E. A., & Mayer, J. Anatomical connections between medial and lateral regions of the hypothalamus concerned with food intake. *Science*, 1967, **157**, 1574–1575.

Balagura, S., & Devenport, L. D. Feeding patterns of normal and ventromedial hypothalamic lesioned male and female rats. *Journal of Comparative and Physiological Psychology*, 1970, **71**, 357–364.

Booth, D. A. Amphetamine anorexia by direct action on the adrenergic feeding system of rat hypothalamus. *Nature*, 1968, **217**, 869–870.

Brobeck, J. R. Mechanism of the development of obesity in animals with hypothalamic lesions. *Physiological Review*, 1946, **26**, 541–559.

Brobeck, J. R., Larsson, S., & Reyes, E. A Study of electrical activity of the hypothalamic feeding mechanism. *Journal of Physiology*, 1956, **132**, 358–364.

Brobeck, J. R., Tepperman, J., & Long, C. N. H. Experimental hypothalamic hyperphagia in the albino rat. *Yale Journal of Biology and Medicine*, 1943, **15**, 831–852.

Brooks, C. McC., Lockwood, R. A., & Wiggins, M. L. A Study of the effects of hypothalamic lesions on the eating habits of the albino rat. *American Journal of Physiology*, 1946, **147**, 735–741.

Carey, R. J., & Salim, A. P. Changes in d-amphetamine reactivity resulting from septal forebrain injury. *Physiology and Behavior*, 1970, **5**, 133–136.

Carlisle, H. J. Differential effects of amphetamine on food and water intake in rats with lateral hypothalamic lesions. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 47–54.

Carlisle, H. J., & Reynolds, R. W. Effect of amphetamine on food intake in rats with brain-stem lesions. *American Journal of Physiology*, 1961, **201**, 965–967.

Carlton, P. L. Cholinergic mechanisms in the control of behavior by the brain. *Psychological Review*, 1963, **70**, 19–39.

Carr, L. A., & Moore, K. E. Norepinephrine: Release from brain by d-amphetamine in vivo. *Science*, 1969, **164**, 322–323.

Cole, S. O. Increased suppression of food intake by amphetamine in rats with anterior hypothalamic lesions. *Journal of Comparative and Physiological Psychology*, 1966, **61**, 302–305.

Cole, S. O. Experimental effects of amphetamine: A review. *Psychological Bulletin*, 1967, **68**, 81–90.

Cole, S. O. Brain mechanisms and depressant action of amphetamine on feeding behavior. *Psychological Reports*, 1968, **23**, 775–782.

Cole, S. O. Experimental effects of amphetamine: Supplementary report. *Perceptual and Motor Skills*, 1970, **31**, 223–232.

Cox, V., & Kakolewski, J. W. Sex differences in body weight regulation in rats following lateral hypothalamic lesions. *Communications in Behavioral Biology*, 1970, **5**, 195–197.

Cox, V. C., Kakolewski, J. W., & Valenstein, E. S. Ventromedial hypothalamic lesions and changes in body weight and food consumption in male and female rats. *Journal of Comparative and Physiological Psychology*, 1969, **67**, 320–326.

Delgado, J. W. R., & Anand, B. K. Increase of food intake induced by electrical stimulation of the lateral hypothalamus. *American Journal of Physiology,* 1953, **172**, 162–168.

Devor, M. G., Wise, R. A., Milgram, N. W., & Hoebel, B. G. Physiological control of hypothalamically elicited feeding and drinking. *Journal of Comparative and Physiological Psychology,* 1970, **73**, 226–232.

Di Cara, L. V. Role of postoperative feeding experience in recovery from lateral hypothalamic damage. *Journal of Comparative and Physiological Psychology,* 1970, **72**, 60–65.

Epstein, A. Suppression of eating and drinking by amphetamine and other drugs in normal and hyperphagic rats. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 37–45.

Feldman, S. E., Larsson, S., Dimick, M. K., & Lepkovsky, S. Aphagia in chickens. *American Journal of Physiology,* 1957, **191**, 259–261.

Foxwell, M. H., Funderburk, W. H., & Ward, J. W. Studies on the site of action of a new anorectic agent, fenfluramine. *Journal of Pharmacology and Experimental Therapeutics,* 1969, **165**, 60–70.

Fuller, J. L., Rosvold, H. E., & Pribram, K. H. The effect on affective and cognitive behavior in the dog of lesions of the pyriform-amygdala-hippocampal cortex. *Journal of Comparative and Physiological Psychology,* 1957, **50**, 89–96.

Glick, S. D. Effects of d-amphetamine and frontal ablations on response suppression in rats. *Journal of Comparative and Physiological Psychology,* 1969, **69**, 49–54.

Glick, S. D., & Jarvik, M. E. Differential effects of amphetamine and scopolamine on matching performance of monkeys with lateral frontal lesions. *Journal of Comparative and Physiological Psychology,* 1970, **73**, 307–313.

Gluckman, M. I., & Baum, T. The pharmacology of iprindole, a new anti-depressant. *Psychopharmacologia,* 1969, **15**, 169–185.

Gold, R. M. Hypothalamic hyperphagia produced by parasagittal knife cuts. *Physiology and Behavior,* 1970, **5**, 23–25. (a)

Gold, R. M. Hypothalamic hyperphagia: Males get just as fat as females. *Journal of Comparative and Physiological Psychology,* 1970, **71**, 347–356. (b)

Graff, H., & Stellar, E. Hyperphagia, obesity, and finickiness. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 418–424.

Grossman, S. P. Eating or drinking elicited by direct adrenergic or cholinergic stimulation of the hypothalamus. *Science,* 1960, **132**, 301–302.

Grossman, S. P. Behavioral effects of chemical stimulation of the ventral amygdala. *Journal of Comparative and Physiological Psychology,* 1964, **57**, 29–36.

Hamilton, C. L. Interaction of food intake and temperature regulation in the rat. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 476–488.

Harris, S. C., Ivy, A. C., & Searle, L. M. The mechanism of amphetamine-induced loss of weight. *Journal of the American Medical Association,* 1947, **134**, 1468–1475.

Hetherington, A. W. The production of hypothalamic obesity in rats already displaying chronic hypopituitarism. *American Journal of Physiology,* 1943, **140**, 89–92.

Hetherington, A. W. Non-production of hypothalamic obesity in the rat by lesions rostral to the ventro-medial hypothalamic nuclei. *Journal of Comparative Neurology,* 1944, **80**, 33–45.

Hoebel, B. G. Hypothalamic lesions by electrocauterization: Disinhibition of feeding and self-stimulation. *Science,* 1965, **149**, 452–453.

Hoebel, B. G. Feeding: Neural control of intake. *Annual Review of Physiology,* 1971, **33**, 533–568.

Hoebel, B. G., & Teitelbaum, P. Hypothalamic control of feeding and self-stimulation. *Science,* 1962, **135**, 375–376.

Jonason, K. R., Lauber, S. M., Robbins, M. J., Meyer, P. M., & Meyer, D. R. Effects of amphetamine upon relearning pattern and black-white discriminations following neocortical lesions in rats. *Journal of Comparative and Physiological Psychology,* 1970, **73**, 47–55.

Khavari, A. Effects of central versus intraperitoneal d-amphetamine administration on learned behavior. *Journal of Comparative and Physiological Psychology,* 1969, **68**, 226–234.

Kirschbaum, W. R. Excessive hunger as a symptom of cerebral origin. *Journal of Nervous and Mental Disease,* 1951, **113**, 94–114.

Krebs, H., Bindra, D., & Campbell, J. F. Effects of amphetamine on neural activity in the hypothalamus. *Physiology and Behavior,* 1969, **4**, 685–691.

Leibowitz, S. F. Hypothalamic β-adrenergic "satiety" system antagonizes an α-adrenergic "hunger" system in the rat. *Nature,* 1970, **226**, 963–964.

Liebelt, R. A., & Perry, J. H. Hypothalamic lesions assosciated with goldthioglucose-induced obesity. *Proceedings of Society for Experimental Biology and Medicine,* 1957, **95**, 774–777.

Maickel, R. P., Cox, R. H., Jr., Miller, F. P., Segal, D., & Russell, R. W. Correlation of brain levels of drugs with behavioral effects. *Journal of Pharmacology and Experimental Therapeutics,* 1969, **165**, 216–224.

Margules, D. L. Noradrenergic synapses in perifornical hypothalamus for the suppression of feeding behavior by satiety. *Federation Proceedings,* 1969, **28**, 641. (Abstract)

Margules, D. L. Alpha-adrenergic receptors in hypothalamus for the suppression of feeding behavior by satiety. *Journal of Comparative and Physiological Psychology,* 1970, **73**, 1–12.

Marshall, N. B., Barrett, R. J., & Mayer, J. Hypothalamic lesions in goldthioglucose injected mice. *Proceedings of Society for Experimental Biology and Medicine,* 1955, **90**, 240–244.

Meyer, D. R., & Meyer, P. M. Brain functions. *Annual Review of Psychology,* 1963, **14**, 155–174.

Miller, N. E. Motivational effects of brain stimulation and drugs. *Federation Proceedings,* 1960, **19**, 846–854.

Mogenson, G. J. Effects of amphetamine on self-stimulation and induced drinking. *Physiology and Behavior,* 1968, **3**, 133–136.

Morgane, P. J. Distinct feeding and hunger motivating systems in the lateral hypothalamus of the rat. *Science,* 1961, **133**, 887–888. (a)

Morgane, P. J. Electrophysiological studies of feeding and satiety centers in the rat. *American Journal of Physiology,* 1961, **201**, 838–844. (b)

Morrison, S. D. The relationship of energy expenditure and spontaneous activity to the aphagia of rats with lesions in the lateral hypothalamus. *Journal of Physiology,* 1968, **197**, 325–343.

Murphy, H. M., & Brown, T. S. Effects of hippocampal lesions on simple and preferential consummatory behavior in the rat. *Journal of Comparative and Physiological Psychology,* 1970, **72**, 404–415.

Nathanson, M. H. The central action of benzedrine. *Journal of the American Medical Association,* 1937, **108**, 528–531.

Novick, I., & Pihl, R. Effect of amphetamine on the septal syndrome in rats. *Journal of Comparative and Physiological Psychology,* 1969, **68**, 220–225.

Olds, J. Hypothalamic substrates of reward. *Physiological Review,* 1962, **42**, 554–604.

Pickens, R., & Harris, W. C. Self-administration of d-amphetamine by rats. *Psychopharmacologia,* 1968, **12**, 158–163.

Poschel, B. P. H., & Ninteman, F. W. Hypothalamic self-stimulation: Its suppression by blockade of norepinephrine synthesis and reinstatement by methamphetamine. *Life Sciences,* 1966, **5**, 11–16.

Powley, T. L., & Keesey, R. E. Relationship of body weight to the lateral hypothalamic feeding syndrome. *Journal of Comparative and Physiological Psychology,* 1970, **70**, 25–36.

Pribram, K. H., & Bagshaw, M. Further analysis of the temporal lobe syndrome utilizing frontotemporal ablations. *Journal of Comparative Neurology,* 1953, **99**, 347–375.

Reiter, L. Effects of amphetamine on lateral hypothalamic activity in response to amygdaloid stimulation. *Federation Proceedings,* 1970, **29**, 383. (Abstract)

Reynolds, R. W. The effect of amphetamine on food intake in normal and hypothalamic hyperphagic rats. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 682–684.

Reynolds, R. W. Ventromedial hypothalamic lesions without hyperphagia. *American Journal of Physiology,* 1963, **204**, 60–62.

Reynolds, R. W. An irritative hypothesis concerning the hypothalamic regulation of food intake. *Psychological Review,* 1965, **72**, 105–116.

Rodgers, W. L., Epstein, A. N., & Teitelbaum, P. Lateral hypothalamic aphagia: Motor failure or motivational deficit? *American Journal of Physiology,* 1965, **208**, 334–342.

Ruch, T. C., Patton, H. D., & Brobeck, J. R. Hyperphagia and adiposity in relation to disturbances of taste. *Federation Proceedings,* 1942, **1**, 76. (Abstract)

Sclafani, A., Belluzzi, J. D., & Grossman, S. P. Effects of lesions in the hypothalamus and amygdala on feeding behavior in the rat. *Journal of Comparative and Physiological Psychology,* 1970, **72**, 394–403.

Sharp, J. C., Neilson, H. C., & Porter, P. B. The effect of amphetamine upon cats with lesions in the ventromedial hypothalamus. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 198–200.

Stark, P., & Totty, C. W. Effects of amphetamine on eating elicited by hypothalamic stimulation. *Journal of Pharmacology and Experimental Therapeutics,* 1967, **158**, 272–278.

Stein, L. Amphetamine and neural reward mechanisms. In H. Steinberg (Ed.), *Animals behavior and drug action.* London: J. & A. Churchill, 1964.

Stowe, F. R., & Miller, A. T. The effect of amphetamine on food intake in rats with hypothalamic hyperphagia. *Experientia,* 1957, **13**, 114–115.

Strominger, J. L., Brobeck, J. R., & Cort, R. L. Regulation of food intake in normal rats and in rats with hypothalamic hyperphagia. *Yale Journal of Biology and Medicine,* 1953, **26**, 55–74.

Teitelbaum, P., & Stellar, E. Recovery from the failure to eat produced by hypothalamic lesions. *Science,* 1954, **120**, 894–895.

Teitelbaum, P., & Epstein, A. The lateral hypothalamic syndrome: Recovery of feeding and drinking after lateral hypothalamic lesions. *Psychological Review,* 1962, **69**, 74–90.

Thode, W. F., & Carlisle, H. J. Effect of lateral hypothalamic stimulation on amphetamine-induced anorexia. *Journal of Comparative and Physiological Psychology,* 1968, **66**, 547–548.

Van Rossum, J. M., & Simons, F. Locomotor activity and anorexogenic action. *Psychopharmacologia,* 1969, **14**, 248–254.

Wade, G. N., & Zucker, I. Modulation of food intake and locomotor activity in female rats by diencephalic hormone implants. *Journal of Comparative and Physiological Psychology,* 1970, **72**, 328–336.

Wagner, J. W., & de Groot, J. Effect of goldthioglucose injections on survival organ damage and obesity in the rat. *Proceedings of Society for Experimental Biology and Medicine,* 1963, **112**, 33–37.

Weiss, B., & Laties, V. G. Behavioral pharmacology and toxicology. *Annual Review of Pharmacology,* 1969, **9**, 297–326.

White, N. M., & Fisher, A. E. Relationship between amygdala and hypothalamus in the control of eating behavior. *Physiology and Behavior,* 1969, **4**, 199–205.

Williams, D. R., & Teitelbaum, P. Some observations on the starvation resulting from lateral hypothalamic lesions. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 458–465.

Wyrwicka, W., & Dobrzecka, C. Relationship between feeding and satiation centers of the hypothalamus. *Science,* 1960, **132**, 805–806.

(Received June 25, 1971)

Psychological Bulletin
1973, Vol. 79, No. 1, 21-44

OLFACTORY CONTROL OF BEHAVIOR IN RODENTS[1]

E. FRED SCHULTZ AND JACK T. TAPP[2]
Vanderbilt University

Behavioral literature is examined for information that could clarify the manner in which rodent behavior is mediated by olfaction. Functional use of olfaction is readily apparent in some of the behaviors reviewed, for example, bait shyness, olfactorally modulated aggression, olfactory marking of and recognition of stressful environments, and inhibition of food seeking by odors generated from frustration. Other behaviors are reviewed in which olfaction is thought to play an important, though yet undelineated role, for example, differences in approach preference based on sexual dimensions. Still other olfactory-associated behaviors are reviewed which meet current standards of inferential significance yet lack any apparent function for the animal. It is concluded that a coherent understanding of the functional use of olfaction by rodents awaits a shift in the philosophy of research away from rigorous examination of loosely selected odorants and behavioral measures toward rigorous examination of normal behaviors in natural social contexts.

The use of the sense of smell by laboratory animals provides investigators of animal behavior with a subject somewhat akin to the weather—a topic generally discussed but seldom acted upon. The olfactory capabilities of laboratory animals, particularly the rat, are viewed with a respect bordering on awe. Research is obviously necessary to determine the limits of these capabilities and to establish a factual foundation for such a point of view, but more importantly, a conceptual orientation is necessary to give this research direction. This is particularly true when one views the origins of research on olfaction in the rat. Laboratory studies of rodent olfaction have grown out of rodent control programs, extensions of experimental techniques from other sense systems to olfaction, and attempts to explain aberant experimental results in learning studies. The study of pheromones has also had its effects on rodent research (Gleason & Reyneirse, 1969), as have ecological studies of the rat. However, these varied motivations for research have produced a spotty coverage of the range of activities in which odors are suspected to function.

The purpose of this review is to attempt to impose some order on the myriad of studies on the olfactory control of rodent choice behaviors that have appeared in the literature within the past decade. The emphasis of this review is on the functional properties that odors have in the regulation of rodent behaviors. We follow the organization used by Barnett (1963) in his comprehensive study of the behavior of the rat because of the functional emphasis of this work. Within this organization there are repeated references to approach and avoidance behaviors observed by the investigators, but these behaviors gain their significance in the context of the total, functionally defined behavior. The review has been limited to the molar behavior of the macrosmotic order *Rodentia,* and studies of discrimination, sensitivity, or other "trained" responses to odors are generally excluded.

MOVEMENT IN THE LIVING SPACE

For the majority of rodents, odors represent the leading edge of the environment. When the visual world is bordered by the nearest pebble, clump of grass, or baseboard, a nose that is within a couple of inches of the ground offers enormous potential for directing behavior. Deferring discussion of movement associated with feeding, social, and sexual behavior, there remains the body of research on movement which relates to the role of odors in exploration.

---

[1] This review was supported in part by Research Grants NB-08855 and GB-7961 from the National Science Foundation to the junior author.

[2] Requests for reprints should be sent to Jack T. Tapp, Suicide Prevention and Crisis Service, 560 Main St., Buffalo, New York 14202.

Date 6.3.81
J 719

Response to Novel Odors

Like other stimuli, odors can be defined as novel when the test animal has had no previous known experiences with the odor. Experiments examining the responses of rodents to novel odors have employed a fairly wide range of odors off the shelf though the test for responding has typically been simple approach or withdrawal. The rationale for these studies has differed from different laboratories.

Reiff (1956) made the initial assumption that the odors of normal food and odors from natural habitats were made attractive by discrete chemical components in the odor complex. He evaluated the preferences exhibited by wild rodents for a number of chemically pure olfactory stimuli. His sample included black rats (*Rattus rattus*), brown rats (*Rattus norvegicus*), house mice (*Mus musculus*), field mice (*Microtus arvalis*), and voles (*Arvicola terrestris*), as well as white mice and rats. The test odorants were (*a*) oils, essences, and extracts from animal and vegetable tissue and (*b*) organic acids and amines. These were placed in open jars in an open field. During the first phase of experimentation 10 pairs of white mice or white rats were placed in the field, and a record was made of the number of approaches to each substance. Each substance was tested several times in different combinations and Reiff tested "the interesting preparations about 30 times [p. 292]."

The substances receiving further investigation were presented to some combination of the captured wild rodents. From the report of the results it was difficult to discriminate between data obtained from the wild rats and mice and that obtained from the albinos. Each presentation of several odors produced a total number of approaches to the jars, and a percentage of approaches was calculated for each substance. An average was then made of the percentage of approach scores for each substance and these average scores were scaled so that a value of 100 represented indifference with respect to the comparison substances. No blank control was run to provide an absolute reference point and no variance statistics were reported, limiting the inferential properties of the measure. Retesting the "interesting" odors systematically changed the nature of the base line of comparison and resulted in the potential loss of the descriptive statistics ordinal properties. Finally, subjects were used for repeated observations and no mention was made of subject scheduling. Reiff's results must therefore be viewed with reservation, and considering the large number of substances he investigated, this is unfortunate.

Of the odorants which closely resembled natural food and body odors, the following substances scored higher than 100: ambergris; natural and synthetic musk; furfuraldehyde (bran oil); vanillin; oil of peppermint; and oils from cod liver, rape seed, cotton seed, soy beans, thyme, and anise. The following substances scored lower than 100: civet; valerian; sperm oil, and oils from linseed, corn, peanuts, fennel, parsley, and celery.

Albino rats and mice were used to test over 200 organic acids and amines, which were assumed to be a fractional component of naturally attractive mixtures. After comparing results from the two groups of odorants, it was Reiff's opinion that no isolated compound was as effective an attractant as the naturally occurring mixtures. Contrary to Reiff's initial hypothesis, it appeared that the odors from normal food and from natural habitats are not made attractive by their discrete chemical components. He did observe high preference values for dihydroxyethylaniline, dihydroxyethylbenzylamine, and diisobutylbenzylamine and concluded that tertiary amines were effective attractants. It is unfortunate that Reiff's results must be viewed with reservation because of the problems implicit in his methodology and in the report of his experiment. The work is of significant value and more investigations should be conducted to follow up the results.

Long and Tapp have examined rat responses to novel odors in a series of studies conceived to evaluate the reinforcing properties of odors. The measure they employed provides some indication of the function of olfactory cues in exploration. Male albino rats were given free-operant tests in a double-lever Skinner box (Long & Tapp, 1968). Pressing either lever delivered a light stream of air in the vicinity of the nose of the animal; one stream was odorized while the other was not. A third air source flushed the chamber of stray odors. Preference was inferred from differences in rates of lever pressing.

In one experiment (Long & Tapp, 1967), one of the odors was obtained by passing air over amyl acetate. The lever that produced the onset of odorized air was pressed significantly more times than the other lever, but this effect decreased over three 30-minute testing periods. This effect was not different in kind from other studies of light-onset reinforcement conducted in the same laboratory (see Tapp, 1969, for review). When a novel odor was compared with light onset and with the onset of other stimuli, the odor was overall more reinforcing, but not qualitatively different from light onset (Tapp & Long, 1968).

The double-lever box was used to evaluate response to a series of novel odorants (Tapp, Plemons, & Spiegler, 1967). The particular odorants were chosen to provide a stimulus from each of the seven categories which Amoore (1952) suggested as representing the primary odors. Two methods of testing were used to determine if both methods would reflect the same heirarchy of odor preferences. In the first study, a single odor was compared with no odor in different groups of animals. In the second study, odors were compared with one another and with no odor. Only three of the substances tested yielded preference values in the same direction by both methods: d-camphor was consistently preferred, while musk ambrette and formaldehyde were consistently not preferred. In all of these experiments, no durable preferences for these odors were observed. If these odors have reinforcing properties, their ability to direct lever pressing is at best transitory.

In other studies conducted by Long and others, more durable reinforcing properties were observed for the odor of food. Though a somewhat different picture emerges for food odor, the studies shed some light on factors that are important in the control of exploration by olfactory stimuli.

In one of these experiments (Long & Tapp, 1967) the odor was obtained by passing air over dry powdered rat chow. There was a higher rate of responding on the lever that produced the food odor than on the control lever (air only) when the test animals were 22-hour food deprived. There was no differential responding for food odor when the rats were satiated.

Long and Stein (1969) demonstrated that the lever-pressing task evaluated not only the attractive component of the odor, but that the response measure was sensitive to the change in stimulus level produced by the response. Three groups of male rats were tested in the double-lever box, and for all groups, the air in one of the carrier airstreams was odorized. For the first group, pressing a lever resulted in the onset of the adjacent airstream. For the second and third groups, both airstreams were on constantly and lever-pressing by the second group turned off the adjacent airstream for 1 second, while lever-pressing by the third group had no effect on the airstreams. All three groups exhibited more responding on the lever adjacent to the odorized airstream than on the one near the unodorized airstream. This rate diminished over five experimental sessions.

When a response produced odor onset, the animals showed the greatest preference for the odorized lever. Turning the air off yielded somewhat less preference for the odor lever. The noncontingent arrangement resulted in the least magnitude of preference for the odorized lever, although the difference between the odorized and the unodorized lever was still significant.

The conclusions of this experiment have direct bearing on the role of olfaction in regulating exploration. The presentation of the odorized air to the noncontingent group attracted the animals to the vicinity of the lever, and they exhibited a higher response rate on the lever nearest the odor. The probability of responding is enhanced if the stimulus changes following a lever press, though the onset of the odor of food is a more powerful reinforcer for lever-pressing than food-odor offset. The stimulus in these latter studies is a highly conditioned reinforcer and the animal has had an extensive history with the odor. For this reason the durability of the reinforcing properties of the food odor was much greater than that observed for any novel odor, and deprivation seems to be a necessary condition for the food odor to have these behavior-directing properties.

The conclusions from this series of experiments are rather clear: (*a*) Like other stimuli, some novel odors can be important stimuli for directing the course of exploratory behaviors, but they lose their behavior-directing properties as they lose their novelty. (*b*) The power of an odor to attract appears to be a function of the animal's past history with the odor. (*c*) The behavior-directing properties of a significant odor such as a food odor can be enhanced when the presence of the odor is contingent on a particular response and the animal is deprived.

Response to Biological Odors

Biological odors in this context are those generated by animals. Their existence is generally inferred from observing the behavior of the test animal. The range of such odors to which a rodent can respond is potentially vast but only a limited number of odors of this sort have yet been examined in detail.

Odors from Conspecific

Efforts to understand the manner in which odors from conspecifics can influence movement about the living space are considerably frustrated by man's inability (or unwillingness) to smell objects and animals in the environment in the same fashion as do the animals themselves. There are four major lines of inquiry which suggest that odors from conspecifics have an influence on movement patterns: territorial marking, spontaneous alternation, stress odors, and frustration odors. It is possible that the odors mediating these four phenomena may not be entirely different, though such a suggestion is currently untested. The existence of other such phenomena is also a very real possibility.

Territorial marking. Territorial marking, along with food selection and mate identification, is one of the most widely acknowledged olfactory functions that urban man ascribes to mammals. Convincing research on territorial marking, particularly in rodents, is conspicuously lacking. A large number of investigators have observed behavior resembling the marking of territory with odorants, and others have observed behavior that seems consistent with the existence of territorial marks, but much remains to be tied together. (See Eisenberg & Klein, 1972 and Ralls, 1971.)

Some behavior patterns suggestive of territorial marking have been studied in rodents. Reiff (1956) found that the faintly visible marks left by rats in their communal trails and in nests possessed an odor which elicited investigation if used as a test odorant. He found that rats' urine had the same effect if it were allowed to ferment, but this attractant property was not observed if the urine were fresh or if the solids, including the bacteria, were removed by centrifugation before aging.

A number of species of rodents possess externally ducted glands and indulge in rubbing behavior that is frequently referred to as marking. Among rodents for which such behaviors have been studied are members of two families: (*a*) *Heteromyidae,* which is the family of pocket mice (Eisenberg, 1963) and (*b*) *Cricetidae,* which includes both gerbils (Baran & Glickman, 1970; Thiessen, Blume, & Lindzey, 1970) and new world rats and mice. Although the evidence suggests that other conspecifics respond differentially to objects if they are marked or unmarked, the dimensions of the behavior observed do not yet suggest a territorial delineating function. Additional study is indicated.

Spontaneous alternation. Odors from conspecifics have effects on the regulation of exploration in a novel environment. Specifically, it is rather clear from a number of studies that the rat's presence in a novel environment can provide cues that decrease the chances that the animal will return to that same place at the next opportunity. One of the major cues that produces this behavior pattern appears to be odor left by the animal.

The identification of this function of odor has arisen from research on spontaneous alternation, a behavior pattern in which an animal enters an alley different from one chosen previously during the unrewarded exploration of a T-maze or a cross-maze.

Douglas (1966) examined the theoretical effect of stimulus preference upon choice behavior and derived a formula that parceled from the observed alternation rate the effect of both chance and differential stimulus attractiveness. This formula yielded a "true" alternation tendency. He empirically validated the assumed statistical independence of repeated-alternation measures on a homogeneous population of animals and derived a parametric base line for further investigations of alternation behavior. The experimental problem was to examine the role of intramaze, extramaze, and response-induced cues.

The results of his studies showed that animals deviated from chance alternation for two reasons only: (*a*) to avoid their own odor trail and (*b*) to walk in a direction 180° from the direction chosen previously. These effects were demonstrated by providing a fresh paper floor, which eliminated alternation to an odor trail, and by passively turning the animals for eight rapid horizontal arcs between trials to disrupt the ability to alternate based on a vestibular-maintained spatial map. Rats did not alternate to the extramaze olfactory cues provided by pepperminty-smelling liquid soap at the end of one arm and by aromatic pipe tobacco at the other. They did show a preference for the pepperminty odor indicating that it was noticed. The existence of a spatial map provided a higher percentage of alternation than did odor trail avoidance, and the additive effect of the two cues combined accounted for the alternation rate when all cues were present.

Brill (1967) demonstrated that male rats spontaneously alternated when they were allowed to use either an odor trail as a cue, or a spatial map based on the vestibular senses as a cue, or both. Alternation was eliminated on test trials when both the paper floor was replaced and the maze was rotated 90°. This result was confirmed in two experiments offering substantial support for Douglas's (1966) conclusions.

Rosen (1968) has provided additional support for Douglas's conclusions in an experiment in which the attempted to assess the relative importance of vestibular and olfactory cues. Only one exposure trial and one test trial were given to rats. Bilateral olfactory bulb ablations and cleaning the maze with aerosol Windex reduced odor cues. Both methods eliminated alternation to odor trail cues. Passive horizontal rotation of the subject through 180° eliminated alternation to spatial cues when odor was not available.

A rotatable cross-maze was used to put spatial cues in conflict with the odor trail left in the choice arms. When this was done, alternation to spatial cues occurred with the resulting behavior being approach to the odor trail. Latencies were significantly higher on test trials when either odor or spatial cues were absent or were in conflict.

Still and MacMillan (1969) administered alternating trials involving four odor trail conditions to 24 male and 24 female hooded rats. The vestibular map was disrupted by the use of two T-mazes set at 90° to each other. Manipulation of the floor paper between trials provided four odor trail conditions. In the first condition, the paper was moved to the other maze intact from the first trial to the second. Under this condition, the subjects alternated away from the odor. For the second condition, fresh paper was used for the second trial, and no alternation was observed. These two conditions replicate the findings of Douglas (1966), Brill (1967), and Rosen (1968), with the additional result that females alternated to the same extent as males.

The third condition was created by rotating the floor paper before the second trial so that the termination of the old odor trial underlay the start box and that portion of the trail that had been in the old start box was now in the opposite goal box. To turn away from the odor trail at the choice point would be to repeat the same direction of turn that occurred in the previous trial. Under this condition, the animals alternated left–right directionality (a response-induced cue) and thereby backtracked over the old odor. The fourth condition consisted of placing the old trail under the stem so that when placed in the start box the animal was standing on the odor from the choice point, and upon reaching the choice point, the odor was that of the old start box. In other words, only half of the total trail was used, that part which led from the start box to the choice point. This group also alternated left–right directionality. Apparently in the absence of both a vestibular map and untampered odor trail, there is alternation using the response-induced cue of left–right directionality.

Klein and Brown (1969) also investigated the role of olfaction in spontaneous alternation. Half of their rats were rendered anosmic by ablation of the olfactory bulbs. All animals received successive tests for alternation in a T-maze and for exploratory behavior in an open field maze. The intact animals alternated at a rate significantly greater than chance while the anosmic animals did not. While exploring the open field, the anosmic animals were considerably more active throughout a session than were the intact ones. The anosmic rats did not show the habituation in open-field activity when tested on three successive days that the intact animals showed. These three differences in patterns of exploration could all arise as a consequence of an anosmic animal being unable to recognize an area as being one that he had already explored.

These studies of spontaneous alternation leave little doubt that odor is a critical cue in producing alternation behavior. The exact part played by odors is unclear, though the results suggest that the odor that the animals responded to is one which is left by the animal, and which functions as a sign to guide future choices. Our understanding would be considerably improved if it were known how well rats can discriminate directionality in an old trail and how long the trail lasts. Furthermore, it would be of interest to know if animals could recall facts about the local environment based on cues from a specific segment of the odor trail. Such abilities may not be unreasonable in macrosmatic animals. Finally, this phenomenon has many of the aspects of territorial marking and perhaps could be further explored in this context.

Stress-generated odors. Several recent experiments have investigated an apparent odor that rodents generate when placed under conditions of imminent or actual physical stress. In a clear demonstration, Valenta and Rigby (1968) have shown that rats could use the change in odor created by electrically shocking another rat to serve as a discriminable stimulus. Food-deprived male rats were trained to press a lever for sucrose water solution at a base-line reward schedule of Variable Interval 1. The air that ventilated the Skinner box passed over a normal rat to provide a standard olfactory background. Fifty milliliters of air drawn from the home cage of other odorant animals was injected into this ventilation stream. If the odor came from an unstressed rat, the availability of reward was increased by shifting the schedule to a Fixed Ratio 2 schedule; if the injected odor was from an animal that had just received an electric shock, the subject was subjected to lever-press-contingent footshock delivered on a Fixed Ratio 1 schedule. The dependent variable was the latency to the first lever press following injection of the odorized air. Each of the six animals showed a significantly longer latency when the air was from the stressed rat than when the air was from a nonstressed animal (means of 8.5 seconds for the stressed odor and 1.6 seconds for the nonstressed odor). Although no function was implied, this experiment adequately demonstrated the existence of at least one stress-generated odor that is detectable by conspecifics.

Muller-Velten (1966) described the response of mice to the odor from other mice which had been handled roughly, a response he called an alarm reaction or *angstgeruch*. By an arrangement of one-way doors, a mouse (both *Mus mus musculus* and *Mus mus domesticus* were used) housed in the test apparatus was forced to negotiate a long passageway from the living chamber to the food chamber. A portion of that passageway was divided to provide two parallel and virtually identical paths (choice arms) to the food chamber. A flow of compressed odorized air was introduced into the middle of one of the choice arms and unodorized air into the middle of the other. Odor was produced by passing the compressed air over the odorant subject or over other test substances. The air flowed against the direction of animal movement and was exhausted from the choice point. The odorant animal was made fearful by picking it up and blowing on it. In addition to testing odors from the whole animal, the animal's urine, feces, and footpad secretions were also tested.

Tests with nonfrightened odorant animals resulted in the odorized choice arm being significantly preferred. When the odorant animal was frightened, the odorized pathway was avoided initially but after adaptation by the test animal, it became the preferred pathway. The alarm scent was found to be contained in the urine and not the feces or the footpad secretions. It was most effective for 8 hours and became ineffective by 24 hours. Urine collected from the population as a whole produced neither attraction nor aversion.

A third study detailing the use of olfaction to communicate the existence of stress in the environment has been conducted by King (1969). Male albino rats received 35 pairings of light onset with an unavoidable footshock while restricted in the start half of a one-way shuttle box. Following specific delays and specific treatments to the shuttle box, they were tested 25 times for escape latency with the safe half of the box accessible. Different groups of subjects were tested for escape tendencies after delays from offset of training of 1, 2, 6, 12, and 24 hours. The second manipulation consisted of three different odorous conditions of the box during testing: (*a*) sealed and untouched since the subject's last training trial; (*b*) cleaned immediately after its last use by disassembling it, then hosing with hot water, and drying with hot air; or (*c*) cleaned immediately and later used for conditioning of another rat within the hour immediately prior to the escape test.

All animals tested in a clean shuttle box exhibited extremely slow escape speeds at all delay intervals. These times were not different from the speeds that would have resulted if no escape attempts were made within 60 seconds (the default value). In contrast, the group tested in the presence of odor from a rat that had recently been conditioned showed high escape speeds at 1-hour delay but their escape speeds at all other delay intervals was not different from default base-line.

The group which was tested while the box contained the odor from their own conditioning training produced a U-shaped escape-speed function. Speed at 1-, 12-, and 24-hour delays were of the same high magnitude as the fresh-odor group and were different from the base lines observed in the clean boxes. Speeds at 2- and 6-hour delays were also similar to the fresh-odor group in that they were low and not different from base line.

A second experiment examined escape speeds at long delays in the presence of appropriately old but unfamiliar training odors. The escape speeds were high and did not differ significantly from those of the equivalent group tested in the presence of their own old odor. King points out that the results obtained with appropriately aged odors duplicated the Kamin effect (Kamin, 1963), that is, "retention of a fear-motivated avoidance response is a U-shaped function of postcondition delay [King, 1969, p. 594]." Inappropriately fresh fear odor disrupted that effect at 12 and 24 hours, while an apparatus cleaned of odor sources from fear-conditioned animals did not elicit an escape response to the conditioned stimulus at all.

These findings of King's seem to indicate the existence of a system for the olfactory marking of dangerous places. A more thorough understanding of the functional significance of this phenomenon would be enhanced by two additional types of experiments to test (*a*) the effect on avoidance performance caused by removal of a well-established artificial training odor that had previously served to identify the training chamber and (*b*) experiments to sort out the nature of the conditioned-stimulus–odor pairing, for example, would a preconditioned flashing-light conditioned stimulus and the appropriately aged fear odor evoke escape response from an otherwise novel chamber?

It is reasonable to believe that rodents can communicate specific information about stress by olfaction. Valenta and Rigby (1968) indicated that stress induces an odor which is detectable by other rats. Muller-Valten (1966) demonstrated a withdrawal from the odor of a stressed conspecific. King (1969) appears to have found olfaction used in marking and the later recognition of dangerous locations. At the present time, there is no data to indicate whether the same odor was being generated in all three experiments or whether each experiment tapped one of a large population of olfactory signals.

Frustration-generated odors. McHose and Ludvigson (1966) observed that animals tested late in a runway study behaved in a fashion appropriate to the reinforcement schedules of subjects that had been in the runway at an earlier time. This accidental finding prompted a follow-up study by Ludvigson and Sytsma (1967). They composed a double-alternation running schedule in a runway for two groups of female rats. The first group was run in blocks so that all members consistently experienced a rewarded trial, or they consistently experienced an unrewarded trial. The second group was run in blocks so that half of all the preceding trials were rewarded and half were unrewarded. Speed of running in the middle 1-foot section of the goal box was reported.

The experiment demonstrated that rats from the group that ran in consistently rewarded blocks or unrewarded blocks ran faster for rewarded trials (about 2 feet per second) and slower for nonrewarded trials (less than 1 foot per second). For the group of animals that as a group was neither consistently rewarded nor frustrated, there was no difference in speeds between rewarded and unrewarded trials. Furthermore, treatment differences showed up in extinction suggesting that training served to direct the animal's attention to stimuli that announced a change in the rules of the game. The group trained with reward of their predecessors consistent with their own reward state extinguished almost immediately, whereas the other group persisted in maintaining high running speeds through all eight extinction trials. Both speed separation and extinction differences were reliable.

Amsel, Hug, and Surridge (1969) have also examined the running speeds of rats in the presence of odors resulting from consistent reward or nonreward. Male albino rats were in a runway for one trial per day but were rewarded on alternate days. On any given day all animals received rewarded trials or all received nonrewarded trials. After 72 days the goal-box running speed of the subjects on the unrewarded days had fallen to less than half of the speed shown on rewarded days. The subjects responded faster on 81% of the rewarded trials during the last part of training.

These studies have demonstrated the potential existence of a rat-generated cue that can communicate information to another rat regarding the presence or absence of reward while the animal is physically distant from the reward itself. The phenomena is startling, and the implication for learning studies in the runway are quite obvious. At this point it would appear that the animals were responding to a chemical cue that accompanied frustration from nonreward.

In an attempt to examine the frustration odor phenomenon in more detail Davis and Ludvigson (1969) trained 40 female rats with one trial per day for 30 days to run down a runway for 12 food pellets. The subjects were divided into four groups, created by applying one treatment from each of the following pairs: either reducing the reward to one food pellet to induce frustration or not changing the reward magnitude, and by either swabbing out the interior of the runway after every trial or cleaning the runway only prior to running the first subject from the group. Reward shift produced a decrement in running speed, irrespective of whether the apparatus was cleaned between subjects. This experiment failed to support Ludvigson's earlier demonstration of a frustration odor cue which affected subsequent behavior.

Seago, Ludvigson, and Remley (1970) rendered some of their rat subjects anosmic by removal of the olfactory bulbs, then tested them in the double-alternation paradigm. The intact animals showed the same patterning when run in squads as had been reported by Ludvigson and Sytsma (1967), that is, fast running on rewarded trials and slow running on nonrewarded trials. The anosmic rats ran with uniform speed on all trials, though they could pattern their running speed when a visual cue indicated the presence of reward. Intact rats that were run early showed little difference in speed between rewarded and unrewarded trials. Animals run later were fast on rewarded trials and slow on nonrewarded trials indicating that the cue was cumulative. The speed of the anosmic animals run with the visual cue was not dependent on the number of previous animals. This data suggests that an olfactory cue can communicate the existence of reward or frustration.

Wasserman and Jensen (1969) gave 20 male rats four trials per day for five days in a runway for one 97-milligram food pellet. Ten of the animals were assigned to the odor-producing group, half of them receiving two training trials per day (reward) and the other half receiving two extinction trials per day (frustration). The remaining animals were tested with four rewarded trials per day. The first trial each day was run on fresh paper, while the remaining three trials of the day were run on a counterbalanced sequence of fresh paper, paper used only by a rewarded odorant rat, or paper used by a frustrated odorant rat.

The speed of rewarded animals that ran on flooring used by frustrated animals was significantly slower than when they ran on either new flooring or flooring used by rewarded animals. In this study slowing was induced in animals that were receiving consistent positive rewards.

The experiments of Ludvigson and his colleagues and of Amsel et al. (1969) were investigating additional slowing in animals that were experiencing shifts in reinforcement consistent with the hypothesized odor. This gives reason to believe that there may be an odor of frustration but that (*a*) its effects may not be completely additive with the effects of reward shifts and/or (*b*) its effects may be slight in comparison with the effects of reward shifts.

In still another investigation of the emotion-cued odor, Morrison and Ludvigson (1970) attempted to generate odors of both reward and frustration to be used as discriminable stimuli in a T-maze. The rats which served as odor sources were adapted to eating in the choice point of a T-maze. They generated the hypothesized cues when they found food in the food cup (reward) or failed to find it (frustration). A third cue condition (clean) was produced by replacing the paper on the maze floor with fresh paper.

The experimental animals were divided into four subgroups. One group received odor of reward as a cue for turning into one of the arms of the maze and odor of frustration as a cue for an opposite turn. A second group received odor of reward as one cue and clean floor paper as the other. A third group received odor of frustration and a clean floor as cues. A fourth group received clean flooring on all trials as a control for any extraneous discriminable cues not produced by the experimental manipulations.

Discrimination was achieved within 84 trials by those two groups that had the odor of frustration as a cue, and was not achieved by the two groups that lacked the odor of frustration. Discrimination was not perfect but represented a steady increase over trials and was significantly different from the 50% average of the other two groups.

It appears that a cue generated by nonreward exists and can be used as a discriminable stimulus. Odor generated by reward did not serve as a discriminable stimulus in this test. This conclusion is not only consistent with all of the other experiments in this area but is supported by the Wasserman and Jensen (1969) finding that running speed is slowed by frustration odor but no speed changes occur for the group with clean paper and the group with reward odor.

In summary, there appears to be a discernible cue generated by rats when their expectation of food reward is frustrated. The cue can affect motivated behavior by slowing running speeds to food reward even though the test subject never experiences the frustration himself. The existence of a reward cue has not been supported to date. It appears highly probable that the frustration cue is an odor, but a clear demonstration of the olfactory properties of this cue has not yet been made. For example, it is possible that the cue being used is not a unique odor but is a quantitative difference in urination or defecation.

Odors from Other Species

The influence on the movement of rodents by odors from other species is another under-researched topic. Such odors could be used to mediate defensive or avoidance behaviors or could function in symbiotic or parasitic life-styles.

A freezing behavior shown by some rats in the presence of cats was described by Griffith (1920). He reported that the animals remained stationary from 3 to 30 minutes. The response was sufficiently prepotent to disrupt feeding, maternal care, sleeping, and sexual behavior. A cat presented when sealed in a glass container did not elicit the freezing, nor did a normal cat elicit freezing from rats whose nostrils were plugged with cotton. Curti (1935) examined the same phenomenon on a population of wild and tame rats and mice which showed a smaller incidence of freezing under any circumstances than Griffith found. Curti's high variability and low number of subjects tended to obscure her results, but the data suggested that a glass barrier was generally sufficient to prevent freezing.

Reiff (1956) evaluated investigatory behavior by rats and mice in response to urines collected on farms, in zoos, and in the laboratory. Urine from domestic animals, in particular, the dog, cat, pig, cow, and horse, were generally attractive when tested after 2 weeks of aging. The same was true of urine from other rats. Only one fresh urine tested proved to be attractive (from the puma). The meaning of these results in unclear at present.

A third study which ingeniously investigated the response of rats to familiar and unfamiliar odors from nonconspecifics was done by McCall, Lester, and Dolan (1969). They examined exploration patterns of seven rats as influenced by the odor of two humans. Young female rats were divided between two caretakers who provided all maintenance for their own animals, including handling once a week for 11 weeks. The animals were then tested for side preference while in an open-field maze. Play objects and caretaker position were systematically varied about the diagonal such that, for any particular subject, her own caretaker was standing outside one off-diagonal corner and the strange caretaker was standing outside the other. The animals were tested twice, once with no lid on the open field, and then again using a transparent, airtight lid for half of the subjects, and an opaque but olfactorally permeable lid for the other half.

The animals spent significantly more time on the side of their own caretaker under those conditions when they could smell him. More preference was shown for the experimenter had worn the same unlaundered sweat shirt for the entire experiment than for the one who had worn a clean T-shirt every day (McCall, Lester, & Corter, 1969). Although the results could be explained by either a preference or an aversion hypothesis, they clearly demonstrate the existence of a behavioral influence (and potential bias) which had previously been unrecognized.

The work of McCall et al. (1969) on human odor influence, along with the work of Valenta and Rigby (1968), Müller-Velten (1966), and King (1969) on odor and stress, and the work of the 12 investigators who have recently written about frustration and its odor, deserve special attention. Not only what was found, but also the fact that the phenomena were discovered at all, is important for a number of reasons. First, all of the findings were unpredicted. That is to say, there exist no directly comparable phenomena in human olfactory experience, and as a result, this use of odor by macrosmates was not widely anticipated. The fact that the phenomena seem reasonable in retrospect should provide no comfort since hindsight does little to expand the frontiers of knowledge.

Second, and more important, these findings and others begin to suggest an extensive reassessment of current experimental methodology for all behavioral research using animals. To say that odor exists as simply another cue to be controlled grossly understates the situation. Odor removal techniques will have to be vastly improved from the current state of the art. Methodology will be further changed by the necessity for more compulsive techniques involving animal handling and running. Investigators will be justified in acting as though spirits were assisting the conventional senses of their subjects.

FEEDING BEHAVIOR

The regulation of movement in living space is also controlled by the animal's nutritional requirements. In this context there has been a small literature which has examined some of the details relating to the capacity of an odor to influence the rat's movements toward, or choice of, foods.

There is a general acceptance of the proposition that the olfactory system of macrosmatic animals differs from that of man in terms of (*a*) the range of odorants to which the system is responsive, and (*b*) the quantal level of sensitivity. In a demonstration of the extended range of olfactory sensitivity, Miller and Erickson (1966) have shown that rats could discriminate between pure water and water containing various chloride salts (sodium, potassium, lithium, and ammonium) using an odor. Four rats (one for each salt) were trained to sniff at two stimulus tubes and then to push through the one-way door adjacent to the tube containing pure water to receive 10-second access to a water reward. All subjects performed reliably at salt concentrations as low as .01 molar. Sectioning the olfactory nerves caudal to the cribiform plate in one subject reduced his performance level to chance. The olfactory sensitivity demonstrated in this study supports the assertion that rodents possess a range of olfactory capabilities unsuspected by humans.

The other frequently ignored dimension of olfactory capability in rodents is absolute quantal sensitivity. Several investigators have tried to measure absolute threshold for pure chemical odorants (Goff, 1961; Moulton, 1960); however, a serendipitous finding by Southall and Long (1969) best serves to put in perspective the rat's sensitivity to food odors. They found that the odor from one dry 45 milligram Noyes food pellet provided sufficient cue value for correct T-maze performance.

Male rats were trained to discriminate amyl acetate from butyric acid in an enclosed T-maze, modified to provide constant airflow from the goal boxes through the choice point to the start box. Concentration of the odorants was reduced until they were eliminated, yet this did not disrupt correct performance. Only when the food reward was not in place at the time of running did discrimination break down. One pellet served to maintain correct responding in most of the animals. In a subsequent experiment (Southall, 1969) faster acquisition occurred using six pellets and no novel odorants.

These procedures differed from standard T-maze experiment in two important cue dimensions: (*a*) The forced airflow eliminated information (and/or motivation) arising from the concentration gradient which surrounds an odorant in still air; (*b*) the airflow provided a mechanical gradient to which the subject could orient. If an airflow-trained subject had been tested under the no-airflow condition and had been able to perform successfully, havoc would have been wrought upon the whole literature of food-reward T maze studies.

Approach Behavior

The research dealing with approach and feeding behavior is of two types: those studies that have probed the attraction that food odors have for rodents and those that have centered on preferences among food odors.

Attraction to the Odor of Food

Sveredenko (1954) investigated the ability of yellow-throated forest mice (*Apodemus flavicollis*) and voles (*Clethrionomys glarelus*) to locate sunflower seeds and hazel nuts that had been buried in the ground. The animals located seed to depths of 30 centimeters, but they reexcavated the path used by the experimenter in burying the seeds, indicating the potential use of other cues in addition to or instead of odor. The animals did not dig up empty burial sites, though they successfully dug for nuts that were ground up to eliminate tactile cues.

Howard, Palmateer, and Marsh (1969) have constructed a sophisticated apparatus that allows three odors to be presented simultaneously in the study of preferences. The floor was covered with 3 inches of pine shavings to conceal nine inverted metal tart pans which served simultaneously as a sensing head for body capacitance relays and as an odor site from which emanated compressed air. Odor was procured by passing the compressed air over jars of food before it was ducted to the sensing heads. Old odors were evacuated via diffusion through the hardware cloth top. Tests were run overnight for 15 hours in an illuminated room. Test with ground squirrels (*Spermophilus beechevi*) yielded an average of 14 minutes digging per animal at an oats site, 2½ minutes at a wheat site, and ½ minute at an odorless control site. This system appears conceptually neat and, excepting the lack of adequate exhaust, well engineered.

The double-bar apparatus of Long and Tapp (1968) described earlier is still another system for measuring the attractant properties of odors. It has the potential for measuring the amount of work (lever pressing) animals will do to gain access to an odor.

These studies have all employed methodologies which reflect some dimension of "liking" for odors by rodents, but, in sum, the conclusion is rather obvious, namely, rodents like food odors.

Preferences among Different Food Odors

Howard, Marsh, and Cole (1968) have examined food-odor preferences in a procedure similar to that used by Sveredenko for examining attraction. Cereals and conifer seeds were buried in open petri dishes under 5 centimeters of peat, and deer mice (*Peromyscus maniculatus*) were tested overnight. Sugar pine and jeffrey pine seeds were 100% detected and eaten; oat groats coated with safflower oil showed less detection but were completely consumed. Barley showed a high rate of detection but low consumption. Empty petri dishes were uncovered 20%–50% of the time, reflecting the base-line response rate above which olfactory cues operated. Wheat and oats were detected at the same rate as the control dishes.

In Reiff's (1956) experiment described earlier, he avoided the consumption confound of the Howard et al. (1968) study by testing isolated odorants. These odorants were probably novel to the subjects, but since they were extracts, essences, and oils from food, and organic acids and amines considered to be components of food, the response to them is relevant to feeding behavior. The conclusion that naturally occurring odorants were more highly preferred than the isolated components of the mixture is probably valid in spite of his methodological problems.

Steinbrecher (1962) performed a large scale investigation involving the attractive properties of many odors, including odors from foods. A cross-maze was set in a colony of captured brown rats (*Rattus norvegicus*) and oats were put into containers on the end of each arm. Under each food container there was a glass jar containing an odor source and a paper bag to catch scattered food. One arm of the maze was always left with oats but without an additional odorant. Pressure detectors allowed a mapping of each rat's exploration pattern, and the sequence in which each odor site was visited provided one measure of odor preference. The relative quantities of oats consumed in the presence of each odor was also measured. Each odorant that was visited first or had the most oats consumed in its presence received a scale value of 30. If an odorant ranked second, the value was 15; the third, 5; and if fourth, 2. Means of both the approach and consumption measures were made when a substance was tested more than once, and a ranking was made of all substances for each measure.

The part of Steinbrecher's experiment that is of interest here is that in which separate foodstuffs were compared. Since the rats ate the food that served as the odorant and since they were allowed to make repeated trials over the course of an evening, the effect of food preference became confounded with odor preference. However, the approach ranking observed for the foods was as follows: pork fat, salad oil, edam cheese, liptauer cheese (all with maximum scores), camembert cheese, limburger cheese, seafish, smoked fish, milk, meat, smoked herring, butter, eggs, carrots (which obtained the mean score), bread, salmon, mainzer cheese, oats (the control standard), potatoes, bacon, romadur cheese, tomatoes, roquefort cheese, and hotcake crumbs with vanilla. No generalities can be extracted from this experiment.

At this point there are numerous questions left to examine with respect to relationships between food odors and attraction. None of these experiments suggest any general relevant dimensions of food-odor attractiveness nor do these studies provide insight into issues such as the role of familiarity versus wired-in predispositions in generating preference, or the origin and function of wired-in food odor preferences. Other potentially fruitful lines of inquiry might investigate the relationship between odor preference and biological state, such as pregnancy, lactation, prehibernation fatness, posthibernation hunger, etc. More research is also necessary to delineate the relationships between food odors and nutritional state.

Withdrawal Behavior

Two situations have been found in which odor can induce withdrawal from feeding behavior. In the first of these, some quantitative or qualitative aspect of the odor is aversive, and the animal shows spontaneous withdrawal. In the second, aversion is learned as in bait shyness.

Spontaneous Withdrawal

Barnett and Spencer (1953) evaluated the abilities of some odors and flavors to repel rats from grain stores. In a captured colony of wild brown rats (*Rattus norvegicus*) which normally ate wheat, odors were either mixed with the wheat or put in tins and inserted into the wheat boxes. Aniseed odor had no effect on consumption but aniseed oil mixed 2.5% with the wheat caused avoidance when unflavored wheat was available. When n-butyric acid was mixed with arachis oil and compared against wheat mixed only with arachis oil, butyric acid significantly lowered consumption. The odor of n-butyl mercaptan alone lowered consumption of whole wheat, as did peppermint oil. These experiments had to be terminated before a full range of odorants could be explored systematically. Little can be concluded from the data that were published.

Millman (1968) found aversion to odors while seeking preferences. Compounds serving as odorants were placed in pure form in a glass collar surrounding a lick tube which contained sucrose solutions; thus, the air in the immediate vicinity of the rat's mouth and nose was very concentrated with odors. A collar around another tube contained one of several less odorous substances chosen to visually resemble the odorant. The odorants were chosen as representative of each of Amoore's (1952) seven categories.

Most odorants (i.e., menthone (minty), d-camphor (camphoraceous), formic acid (pungent), ethylene dichloride (ethereal), and butyric acid (rancid)) caused significantly less than half of the total consumption of sucrose solution to be drunk from the odorized tube. None of the odorants enhanced consumption. There are no conclusions to be drawn from these studies at present. Millman's technique for measuring aversions is interesting, though it is obviously necessary to control for concentrations in future studies of this sort.

Learned Withdrawal

Bait shyness, or toxiphobia, functions to protect organisms from poisoning, and represents a conditioned aversion to specific food. To experimentally produce the effect, the ingestion of a food which is novel in some dimension is followed by a sublethal dose of a poison. Pain and Booth (1968) adapted subjects to daily presentation of 20% solution of d-glucose in addition to their ad-lib water and chow. On the first experimental day, the glucose solution was odorized (and potentially flavored) by the addition of 10 μl. of lavender oil per 100 milliliters of glucose solution. Half of the rats received a sublethal intraperitoneal injection of cyclophosphamide following drinking. The presentation of the glucose solution was continued, but the presence of odor was alternated by days.

Consumption of nonodorized days was similar to that observed prior to the injection phase of the study. During the first two odorized presentations, glucose consumption fell by approximately 50% for poisoned animals while saline injected controls showed no significant pattern. Following 9 days of extinction and 10 additional days when no glucose was presented, the same experiment was run on the same animals beginning with 5 days of habituation. The saline controls received the poison, in a slightly larger dosage, and the odorant used was citral. The results were essentially the same as those observed on the first test. If it can be assumed that the odorants did not flavor the solutions, then Pain and Booth satisfactorily demonstrated that an olfactory cue can mediate the toxiphobic response.

Little attention has been given to the capacities of odors to induce withdrawal from food or other objects. The olfactory component in neophobia, which wild rats show to all new objects in their environment, also deserves detailed study.

The Steinbrecher (1962) study on response to food odors deserves consideration apart from its study of attraction to and repulsion from food odors because of the number of substances tested. Using the elevated crossmaze with the odorized feeding stations described earlier, a very large number of odorants was examined, including grains, processed foods, commercial scents, extracts, organic

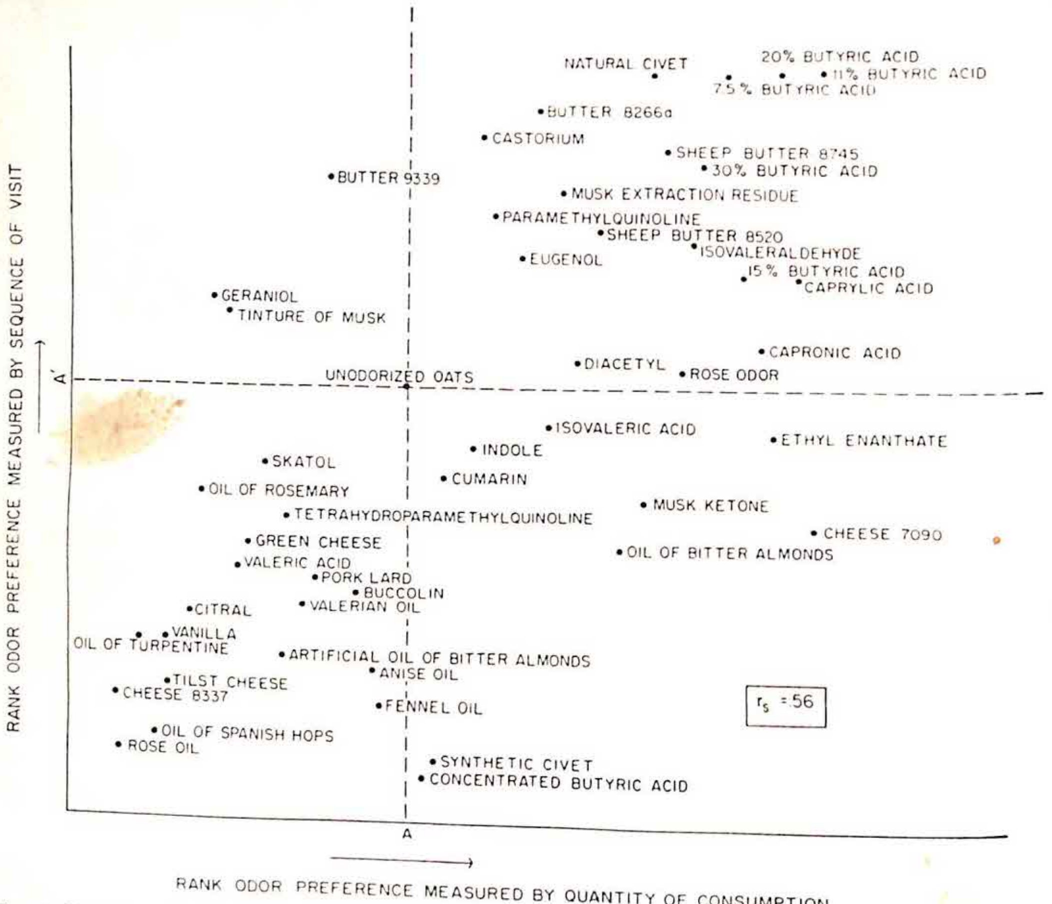


FIG. 1. Ranking of odorants on two preference scales by Steinbrecher (1962). (Points A and A′ are neutral values obtained without the addition of an odorant.)

acids, aromatic hydrocarbons, and esters. An unodorized feeding station was always present. Scales reflecting sequence of approach and the quantity of food consumed were computed. The relative noninterval position of each odorant on both scales is shown in Figure 1. This presentation allows a convenient summary of Steinbrecher's results.

Some odors were not among the first visited but also they elicited greater consumption than the unodorized site. Among these odors were several dilutions of butyric acid, several butter preparations, natural civet, and castorium. Other odors were visited late, well after the unodorized arm, and furthermore, inhibited consumption when compared with the unodorized site. These odors included rose oil, oil of rosemary, vanilla, turpentine, anise oil, oil of Spanish hops, pork lard, citral, skatol, nitrobenzine, and one variety of cheese. More curiously, a few odors such as geraniol and tincture of musk were approached readily but little was eaten in their presence, while other odors, including methyl salicylate, musk ketone, and oil of bitter almonds, were visited well after the unodorized site but elicited much consumption. Concentrated butyric acid always received late visit but elicited the same consumption as the odorless site. Different responses were given to natural odorants and their synthetic counterparts. Real civet was consistently approached first, and synthetic civet last, but their consumption measures were the same. Oil of bitter almonds elicited much consumption, but the synthetic produce saw little consumption. Despite the exceptions, Steinbrecher's measures rank correlate .56 with themselves. Unfortunately, no conclusions other than the empirical description are forthcoming from the study.

RODENT SOCIETIES

Calhoun (1962a) has conducted an extensive observational study of wild brown rats (*Rattus norvegicus*) in the feral state. He found that wild rats have a dynamic social culture that is influenced by the interaction of a large number of factors. The social status of any one individual is some complex function of living site, nearness to food, age, absolute physical size, rate of growth, social rank of the mother, social rank of the mother's colony, age of other animals associated with while a youth, sex ratio of the home colony, and success at fighting. The influence of this social status is seen in many facets of rat life, including style of walking, time of feeding, selection of mates, and reproductive success. Calhoun observed marking, rubbing, and sniffing behavior associated with courtship and reproduction, which strongly suggested that odor was involved. He did not describe a role for odor in other kinds of behavior, and because experimental manipulations disrupted the colony, he did not attempt to extensively investigate the effects of odor.

Other researchers with a more experimental bent have directly investigated odor and social behavior and have found a variety of interesting interactions. Two studies indicate that varying degrees of social traits influence the odors given off by an animal. Several other studies suggest that the reverse is also true—odors influence various dimensions of social behavior.

Influence of Social Traits on Odor

Two experiments suggest that the odors of rats and mice are responded to differently if the odorant animals differ from one another along the dimensions of dominance or aggression. To examine these relationships, Carr and his associates artificially induced dominance or aggression in the odorant animals. Odor stimuli for subsequent tests were created by placing odorant animals into a cardboard test cylinder with a sawdust floor. The ends of the cylinders were closed by permeable partitions, and at one end of each cylinder, the partitition was recessed to form an atrium with a dead-air space. Approach behavior was scored when the subject's head entered any portion of the atrium.

Carr, Martorano, and Krames (1970) trained a group of C57BL/6J mice to be exceedingly savage. Following 6 weeks of isolated living, naive subjects were placed in the home cages of the trained fighters and each subject was attacked and defeated 10 times. Subjects from another group were exposed to the fighters but were separated from them by a wire screen, while subjects from still another group were placed in fighter cages behind a wire screen while the fighter was not present. All subjects were then tested in the vacated home cage of a fighter mouse that contained two test cylinders, one smelling of the fighter mouse that lived in that cage and the other smelling of an unfamiliar defeated mouse. For subjects from one group, the fighter odor was generated by the same fighter mice that had defeated them, but for the remaining three groups, the fighter was from unfamiliar fighters.

Results indicated that all mice spent significantly more time investigating the fighter odor than victim odor unless the subject had been defeated by the particular fighter. In the latter instance, the victim received considerably more exploration time than the fighter. These results provided strong evidence for the existence of an odor complex associated with aggressiveness, and they suggest an interaction between the nature of the odor-complex and the experience of the individual animal in determining the particular behavior pattern observed.

Krames, Carr, and Bergman (1969) observed somewhat different results in a study with rats. A stable dominance hierarchy was established in four groups of rats by providing limited access to food. All subjects were then tested in their own home cage to determine the amount of time spent investigating two odor cylinders smelling of the most dominant rat from an unfamiliar group or smelling of the least dominant rat from the same unfamiliar group. The odor from the least dominant rat received the most investigation.

Both studies demonstrate that rats and mice respond differentially to the body odors of conspecifics when the odors involved differ only in the dominance role of the source animal. Thus it seems that odor is at least one cue used to communicate social status.

Influence of Odor on Social Dimensions

Odor has also been shown to affect aggressive behavior and to affect behavior related to group cohesion.

Influence of Odor on the Dimension of Aggression

Ropartz (1968) paired off members of a colony of Swiss albino mice and selected those which were judged most aggressive. Those chosen were again paired off with one another for a second round of encounters after each had been sprayed with a commercial perfume. Eight days later the same animals had their olfactory bulbs removed and were again paired off. The initial mean latency to attack was 35 seconds. Masking odors with the perfume raised the latency to 199 seconds, and after bulb ablations no attacks occurred. The mean number of attacks per animal fell from 26 to 9 to 0. Bulbectomized mice uniformly submitted to attacks by intact mice despite their former aggressive nature.

The results of an experiment by Archer (1968) suggest that change in the olfactory stimulus constellation may be sufficient to induce aggression. Mice reared in pairs exhibited virtually no aggression. When the pairs were transferred to new clean cages, aggression increased by a slight but significant amount. When the pairs were transferred to different cages recently vacated by pairs of strange male mice of a different strain, aggression increased markedly.

Mackintosh and Grant (1966) reported an experiment in which aggression in pairs of mice housed together was produced by modifying body odor to four levels of unfamiliarity. The animals from Group 1 were rubbed in the perineal region with a cotton wad and then returned to the home cage. The animals from Group 2 were rubbed in the perineal region by the perineal region of a strange male mouse before observation. The experimental animals from Group 3 were used to rub the perineal region of strange mice and the strange mice were placed with the home membrs. The experimental animals in Group 4 wre also replaced by a strange mouse as in Group 3, but the stranger was rubbed only with a cotton wad and bore no trace of a familiar odor.

The magnitude of olfactory novelty increased from Group 1 to Group 4, as did the magnitude of aggression. Both the duration and frequency of aggression was far higher for Group 2 than for Group 1, and for Group 4 than for Group 3. The odor of a strange mouse added to a familiar mouse increased aggression while odor of a familiar mouse added to a strange mouse reduced aggression. In all but Group 1, the aggression of the home member far exceeded the aggression of the experimental member.

A similar experiment was reported simultaneously in which the perineal regions of both pair members were rubbed, using urine-soaked cotton from strange mice instead of themselves, and resulted in the same high levels of aggression.

When taken together, these studies suggest that the olfactory complex contributes to the elicitation of aggressive behavior in rodents. More specifically, the more novel an odor, be it the odor of the test environment or the odor of a conspecific, the greater the degree of aggressive expression. Such a mechanism could have the following two effects on the regulation of the rodent society: (*a*) It could minimize infighting in an established hierarchy and (*b*) it could prevent intrusion by a novel animal into a well-established colony hierarchy.

Influence of Odor on Group Cohesion

Odor can facilitate group cohesion by serving as the mediator of individual attraction. Le Magnen (1952) reported that a sexually naive male Wistar albino rat would approach another male, a female in diestrus, or a female in estrus rather than approach a place lacking animal odor. These tests were conducted in a small closed metal T-maze with the stimulus animals housed in the ends of the cross arms and separated from the goal box by a louvered metal partition. Air was forced through the stimulus holders, into the goal boxes, and then through the choice point area to the start box.

Muller-Velten (1966) has also found that mice captured from a wild interbreeding population of *Mus mus musculus* and *Mus mus domesticus* chose to use the passageway containing the odor of unstressed mice while exploring or seeking food rather than the unodorized passageway. Fresh group urine did not have attractive properties, although males were attracted to female urine. Some feces, even those released in fear, also attracted. These studies by Le Magnen and

Muller-Velten suggest that some component of the body odor of rats and mice attracts conspecifics when the alternative behavior is restricted and that there is some suggestion of differential sex effects.

An olfactory influence in social cohesion has been found in a series of studies by Marr and others on the effects of early experience. The general procedure consists of odorizing a nesting litter for a specific interval then testing the young at a later date for preferences between an unodorized animal, an animal scented with a novel odorant, and an animal scented with odorant previously applied to the subject. An open Y-maze with compartments for the stimulus animals was the test apparatus.

Marr and Gardner (1965) rubbed different rat litters with cloths containing methyl salicylate, Yardley's Red Roses Cologne, and for control, odors picked up by a cloth kept on the bottom of the home cage. Mothers were rubbed prior to birth and pups were rubbed once daily from Day 2 until Day 30. In general, the animals preferred the odor of rearing, spending about 28% of their total time near any of the stimulus animals and about 13% of the time near the animal smelling of the odor of rearing.

Marr and Gardner also examined the influences of early olfactory experience on the mating and breeding behavior of the three groups. Breeding tests with members of the same treatment group showed no difference between those animals that had been odorized and normals in either the amount of time the male followed the female or in the latency to the first mount. However, the once-scented males mounted their once-scented mates only half as often as did the subjects from the nonscented group.

A second test of early olfactory experience on the odor preference was reported by Marr and Lilliston (1969). Novel scent was applied during the first, second, or third week after birth. Two odorants were used and tested—acetophenone and ethyl benzoate. Subjects remained significantly longer near the familiarly scented stimulus animal than the unscented or novel-scented animal. Those subjects scented during the first week showed greater odor preference than did the subjects scented later.

Carter and Marr (1970) followed up this indication of a critical period influencing olfactory preference with Guinea pigs (*Cavia porcellus*) that were scented with acetophenone and with ethyl benzoate for 3 weeks. When tested for preference at Day 14 and Day 22 the animals showed preferences for the odor of rearing. Subjects scented for the first 2 weeks and tested a week later still showed the same preference. Subjects scented for periods of 3 and 5 days during the first 3 weeks and tested during the second, third, and seventh week failed to show preference for the familiarly scented stimulus animal and chose instead the unscented one. Comparisons of preference values across groups revealed that increased exposure yielded a monotonically increasing preference for the experimental scent and a montonically decreasing preference for the naturally smelling animal. To the extent that artificial odorization successfully models the development of natural odor preference, natural preference appears to be a function of exposure length. At this time, there is no evidence for a critical period during which exposure must occur for preference to develop.

Schapiro and Salas (1970) have reported that the odor of the mother rat will inhibit activity in infant rats. They suggest that this may reflect an imprinting phenomena related to individual recognition. Though the suggestion is well made, their experiments fail to provide a test of this hypothesis since they tested only the odor from the mother of the test infant.

For rodents, olfaction is a sensory modality with a potentially powerful influence on a variety of social behaviors. Since some of these odors may be produced by glands, since they effect behavior and may be odorless to man, it is tempting to view them as pheromones. A note of caution is in order. There do appear to exist mammalian pheromones which influence the physiological functions of some species (primer pheromones), specifically the reproductive behavior of some mice (Bruce & Parkes, 1960; Dominic, 1966; Lee & Boot, 1956; Whitten, 1966; Whitten, Bronson, & Greenstein, 1968). On the other hand, there has been a conspicuous lack of evidence for a mammalian releaser pheromone. No immediate response to airborne chemicals has been reported that is not readily explained by conventional learning and olfaction. To suggest at this point that some mammalian social behaviors are the result of pheromones, rather than to other attributes of odors, is to invoke an additional complex conceptual framework in the absence of need and in the face of an adequate and far more parsimonious explanation.

FIXED PATTERNS AND HEREDITY

Were releaser pheromones to exist for mammals, they would most likely regulate fixed-action patterns or species-specific behaviors, that is, the specific movement patterns that are exhibited by animals, which take on a stereotypical form. These behaviors consist of a complex pattern of movements that are shared by the majority of the members of a species and are regulated by complex sets of stimuli. Only a small part of the literature on rodents has investigated the roles of odor as critical stimuli in the elicitation and maintenance of such behavior patterns.

The literature on odor and fixed-action patterns has dealt almost exclusively with the role of odors in sexual behavior. A major exception can be found in a series of experiments conducted by Beach and Jaynes (1956a, 1956b, 1956c) in which they investigated maternal retrieving behavior. Ablations of the olfactory bulbs resulted in indiscriminate retrieval of both the mother's own offspring and of strange rat pups. The olfactory bulbs, and thus, presumably, the sense of smell, were not necessary for retrieval to occur and thus olfaction does not control the total behavioral pattern. However, the experiments did demonstrate that the female uses an olfactory cue to identify a characteristic odor that is associated with her pups in the process of retrieving. This suggests that there may be a "family odor" that is specific to any given mother's set of rat pups. The study by Schapiro and Salas (1970) which shows that pups inhibit activity in the presence of their mother's odor is corroborative. The area warrants further investigation.

Odor in Sexual Direction

There have been a rather large number of experiments concerned with sex and odor preferences which examined several factors that could affect an animal's preference for sex odors. These factors have included the sex of the subject, the subject's hormonal readiness for copulation, the hormonal readiness for copulation in the odorant animals, previous sexual experience on the part of the subject, and the familiarity of the subject with the test animals. Discrepant findings in this literature indicate that there may be still other factors that affect the measurement of sex-odor preferences including potential differences in response caused by odors that are generated from different parts of the body and the ever-present difference in sensitivity among measurement techniques.

The protocol on most of these experiments is similar. A subject is typically presented with a two-choice preference task. The odors that are used as stimuli come from two odorant animals in different physiological states. Characteristically the odorant animals used have been a receptive female, a nonreceptive female, an intact male, a castrated male, and a clean-air control in which no animal was presented as a subject. The subjects themselves differ in sex, in degree of sexual experience, and in hormonal state. As a result, there are 80 possible nonredundant combinations made up of one subject animal and two odorant animals. Table 1 consists of a systematic outline of those combinations and a summary of these experiments.

There are differences between experiments in the mode of generation of the odorants. Several of the experiments presented stimulus odors by placing the entire odorant animal or its excrement in a porous container (Carr, Loeb, & Dissinger, 1965; Carr, Loeb, & Wylie, 1966; Carr, Wylie, & Loeb, 1970). Scott and Pfaff (1970) used the odor of urine in test tubes. Pfaff and Pfaffman (1969) used glass bottles containing urine as odor sources. Stern (1970) used cardboard boxes to contain the stimulus animals at the time of testing. Le Magnen (1953), Müller-Velten (1966), and Pranzarone (1969) blew air over the odorant animals. Brill (1967) used paper flooring which had been walked over.

There are also differences between studies in the measurements employed. Seven of the studies included in Table 1 used as a dependent measure the amount of time spent investigating the stimulus (Carr, et al., 1965; Carr, et al., 1966; Carr, et al., 1970; Pfaff

TABLE 1

Sex Odor Preferences Shown by Male and Female Subjects

Sexual status of nonpreferred odorant animal	Hormonal state preferred odorant animal									
	Prepared to copulate[a]					Not prepared to copulate[b]				
	Receptive female	Nonreceptive female	Intact male	Castrated male	No odor	Receptive female	Nonreceptive female	Intact male	Castrated male	No odor
Male naive subject										
Receptive female		(2), (10)			(1)		(2), (5) (7)	(5)		
Nonreceptive female	(2), 5, 8, (10)		(5)		(1)	(2), (5), (7)		(5)		
Intact male	5	(5)		(4)	1	(5)	(5)		(4)	
Castrated male			(4)					(4)		
No odor	(1), 5	(1), 5	5			(7)	(7)			
Male experienced subject										
Receptive female		(8)					(3) (10)			
Nonreceptive female	2, 3, 7, (8), 10				(7)	(3) (10)				
Intact male				(4)					(4)	
Castrated male			(4)					(4)		
No odor	6, 7	6, (7)								
Female naive subject										
Receptive female		4			1		(4)			(1)
Nonreceptive female					(1)	(4)		(5)		(1)
Intact male					(1)		(5)		(2), (5)	(1)
Castrated male			2, 9					(2), (5), (9)		
No odor		(1)	(1)			(1)	(1)	(1)		
Female experienced subject										
Receptive female		(4)					(4)			
Nonreceptive female	(4)					(4)				
Intact male				(2)						
Castrated male			(2)					2		
No odor										

Note.—Preference is indicated by an open number, indifference by a number in parenthesis. Number key: (1) Brill 1967—Hooded rats; (2) Carr, Loeb, & Dissinger, 1965—Long-Evans hooded rats; (3) Carr, Loeb, & Wylie, 1966—Long-Evans hooded rats; (4) Carr, Wylie & Loeb, 1970—Long-Evans hooded rats; (5) LeMagnen, 1952—Wister albino rats; (6) Müller-Velten, 1966—Wild *mus* mice; (7) Pfaff & Pfaffman, 1969—Sprague-Dawley albino rats; (8) Franzarone, 1969—Long-Evans hooded rats; (9) Scott & Pfaff, 1970—Jackson laboratory mice; (10) Stern, 1970—Sprague-Dawley albino rats.

[a] Male subjects were gonadally intact. Females were in natural or drug-induced estrus.

[b] Male subjects were castrated. Females were in natural diestrus or were ovariectomized.

& Pfaffman, 1969; Pranzarone, 1969; Scott & Pfaff, 1970; Stern, 1970). Other experimenters used a dual passageway which was odorized with body odor and recorded the passageway chosen (Brill, 1967; Le Magnen, 1952; Müller-Velten, 1966).

The results that are summarized in Table 1 offer little evidence of order on a large scale. Of the 34 combinations of variables tested, preference was found in only 13 of these combinations by at least one investigator. Only two of the combinations, however, have yielded systematic preference when examined by more than one investigator, and likewise only two combinations have yielded systematic indifference. The remaining combinations receiving multiple test have produced conflicting results. If all of the data are collapsed and the total pattern presented in Table 1 is examined, it appears that the odor of a receptive female is preferred to the odor of a nonreceptive female, which in turn is preferred to the odor of an intact male, which is preferred to the odor of a castrated male. All of these stimuli are preferred to no odor. It should be borne in mind that such a global ranking is made in the absence of much data and is therefore highly tentative.

On a smaller scale, conclusions that have been made by Carr et al. (1970) warrant closer investigation. For male rats both hormonal readiness for copulation and prior sexual experience were jointly necessary in order to observe a preference for estral odor. In contrast both Le Magnen (1952) and Pranzarone (1969) found naive male rats preferring odors from estral females.

Carr et al. (1970) suggest also that either hormonal readiness for copulation or prior sexual experience were independently sufficient to elicit the preferences shown by female rats for the odor of intact males over that from castrates. In contrast, Scott and Pfaff (1970) have found that naive, diestral female mice showed significant preferences for odors from intact males. When Carr et al. (1965) tested such sexually experienced female rats, they failed to find preference for odors from intact males. It would thus appear that hard and fast rules concerning odor preference, sexual experience, and hormonal readiness are somewhat premature.

The individual findings and discrepancies are more enlightening than the collective results. The experiments by Brill (1967) seem to constitute consistent exceptions to the data of Le Magnen (1952) suggesting that the odor of odor trails laid down in a T maze are responded to differently than odors from the total body or odor from excrement.

Measurement differences also may account for some of the discrepant findings outlined in Table 1. For example, Pfaff and Pfaffman (1969) found that if the measure of investigation was the amount of time spent sniffing at the opening of a bottle which contained the odorant material, then normal male mice showed no differential preference between odors from receptive vs. nonreceptive females. If, on the other hand, the measure was of the amount of time which was spent both at the opening and with the nose touching any part of the glass bottle once the opening has been sniffed, then a different pattern of odor preference emerged. Studies by Pranzarone (1969), Stern (1970), and Carr et al. (1965) used measurement procedures which assessed only the amount of time the animal spent at the opening of the odor container, resulting in potential missed effects.

Stern's (1970) results also provide an insight that may account for some inconsistencies. His experiment showed that in order for male rats to show preference for the odor of estral rather than diestral female rats, sexual experience with females is necessary. The required sexual experience did not have to result in ejaculation. Sexual interaction could be stopped after intromission or after mounting and the preference would still be exhibited, although exposure to receptive females through a wire screen was insufficient to allow the generation of preference.

Carr, Krames, and Costanzo (1970) report a series of experiments which further complicates the ability to specify the dimensions of an adequate sexual experience in development of preference for sex odor. The responses of rats tested on preferences for odors from novel versus familiar sex partners was influenced by at least four variables: the sex of the subject, the number of previous sex acts, the current level of sexual satiation, and whether the history of the animal's sexual experience was monogamous or polygamous. Polygamous males with much experience spent more time examining odors from novel than from familiar partners while female rats with extensive polygamous experience showed indifference with respect to the dimension of novelty versus familiarity. Males with extensive monogamous sexual experience failed to show the odor preferences that their polygamous counterparts had exhibited.

Sexual satiation was also varied at the time of testing by placing all subjects with a cooperative partner shortly before testing. Half of these subjects were removed for testing after one ejaculation while the other half copulated to satiation before testing. The number of ejaculations that immediately preceded the test did not influence the odor preferences for those males that had extensive previous sexual experience, nor did it effect the indifference shown by females (all polygamous). When a group of males that had no sexual contact prior to the pretest copulation was tested, the amount of recent experience did influence preference. Those formerly naive males that had copulated to satiation showed the normal odor preferences characteristic of males of extensive polygamous experience, that is, they preferred the novel odor. However, those males whose only sexual experience was a single pretest ejaculation did not show differential preferences.

Some limited conclusions can be generated from these observations: (*a*) Male rats seem to show preferences for sexual odors under more circumstances than do females. Reference to Table 1 shows that, for males, preferences were found in 8 of the 19 combinations that were tested, whereas for females those preferences were found in only 5 of the 15 combinations. If the summary is restricted to subjects that were hormonally prepared for copulation, then 8 of the 11 combinations presented to male subjects elicited preference, while only 3 of 7 combinations tested with females elicited preference. It has been generally speculated that the preference discrepancy between sexes reflects different roles in mating behavior, but such a suggestion remains to be supported. (*b*) Odors generated by different parts of the body are responded to in different fashions. This is suggested by the contradiction between Brill's (1967) odor trail study and Le Magnen's (1952) body odor study. More direct tests of this hypothesis would seem to be called for. (*c*) Different measurement techniques possess different sensitivities. (*d*) It would appear that sexual experience does contribute substantially to preferences involving sex odors but the specifics of this interaction are not currently understood.

The Role of Odor in Sexual Intercourse

The function of odors in sexual behavior has been examined by eliminating the influence of odors via olfactory bulb ablations. All of these experiments have been conducted on male animals, and all but one have used rats. The change in the behavior following ablations has been examined in both naive and experienced males.

In rats, bulb ablations do not eliminatae sexual intercourse (Beach, 1942; Bermant & Taylor, 1969; Heimer & Larsson, 1967). For sexually experienced males, bulb ablations do not prolong the latency to approach a receptive female, mount her, or to achieve intromission (Bermant & Taylor, 1969; Heimer & Larsson, 1967). Bulb ablations in naive animals, however, were shown to prolong latency to intromission (Bermant & Taylor, 1969). Once intromission was achieved, bulb ablations generally prolonged the interval until ejaculation in both the experienced and naive animals (Bermant & Taylor, 1969; Heimer & Larsson, 1967).

The overall slowing effect of bulbectomy has also been reflected in the number of ejaculations achieved during a relatively short testing session. The number of ejaculations necessary to achieve satiation, however, is not affected by bulbectomies.

All of this evidence suggests very strongly that for rats the ability to smell is not necessary to allow sexual behavior, but it does produce an effect through the interaction of prior experience.

In contrast to findings in rats, Murphy and Schneider (1969) found that male golden hamsters with bulb ablations did not copulate at all. They engaged in no social behavior with other hamsters except a minimum sufficient to remove them from disturbing effects of others. As a consequence, bulbectomized male hamsters initiated no courtship or mating behavior.

The clear-cut discrepancy between responses across rodent species precludes any global conclusion regarding the role of odors in effecting sexual intercourse. Odors appear to have some exciting, directing, and maintaining effect on male rats, and odors appear to be entirely necessary for the male hamsters to engage in sexual intercourse. It is clear that odors play a complex role in reproductive behavior in rodents.

The simplistic approach adopted by most of the investigators discussed here stands out in sharp contrast to the complex sex-oriented behaviors described by Calhoun (1962a). Observations made of wild rat colonies in a quarter-acre compound revealed that the appearance of a female in estrus caused a large number of behavioral changes. Packs of males followed her, sniffing and attempting to mount. She was not allowed to feed or drink in peace. If she attempted to escape into her territory or burrow, they would ignore territorial boundaries to follow her. They would station themselves at all entrances to her burrow and even follow her into it if she remained too long. Aggression by unsatiated males was displaced toward other males encountered elsewhere in the compound. In colonies dominated by a single male, that male protected the female from most of her suitors; but for females from colonies with multiple males, and therefore, with no protection, the number of mountings per night was estimated to approach a thousand. When mounting was allowed, it followed a rigid series of stereotyped steps, that is, fixed-action patterns. The large number of new behaviors and shifts in old behaviors that occur when a female is in estrus offer many possibilities for the interplay of odor with sexual behavior. Further research is certainly warranted to elaborate the eliciting and maintaining roles of odor in such behavior.

CURRENT STATUS AND FUTURE DIRECTION

No consistent generalizations have emerged that allow a convenient summary of relationships between olfaction and rodent behaviors. Perhaps this is because research on olfaction and behavior suffers under two very serious constraints. The major constraint is our ignorance of the qualitative dimension of odor. At this time, investigators cannot describe odors in terms of relevant stimulus parameters and it is, therefore, impossible to perform experiments that require knowledge of the exact differences between olfactory stimuli. In the face of this ignorance, a reasonable alternative strategy for research on the role of olfactory stimuli and animal behavior is to choose odors for investigation that are biologically meaningful, that is, those that stand in an obvious (or hypothesized) relationship to some element of an animal's normal behavior repertoire. This strategy enhances the probability of examining odor for which an integrated behavior pattern exists that, in turn, increases the probability of discovering the circumstances of that odor's use and the adaptive advantages which therefore accrue. This aproach can yield a better understanding of the many ways in which odors are used by animals for coping with the problems of life.

A second constraint encountered by investigators who seek insight into the role played by odor in the life of rodents is that the behavior sought may not be a viable part of the animal's repertoire as a result of restrictive rearing conditions in the laboratory. Both Barnett (1963) and Calhoun (1962a, 1962b) vividly illustrate that rats can develop a sophisticated social life with interaction at many levels, and apparently their genetic inheritance provides them with the predispositions for such development. Yet, standard laboratory rearing and housing conditions may well deprive animals of the opportunity to develop these latent abilities. A case has already been made for doing research with biologically meaningful odors. A strategy that would further optimize payoffs would be to use subjects that have had adequate chance to develop a full range of innate behavior.

Despite these constraints, research on the use of olfaction by rodents to regulate their behaviors has had considerable study. In the search for generalizations the studies of spontaneous preferences for odors off the shelf appear at this point to have been fruitless and they do not appear to lead in any specific direction. On the other hand, there have been several interesting findings concerning use of odor by rodents. Rodents appear to be able to give themselves olfactory cues about their environment and perhaps via these cues they communicate the state of the environment to other animals. There is an odor component in trail marking, in spontaneous alternation, reactions to stress and frustration, social dominance and cohesion, and sexual behavior. These are exciting findings in themselves, and it would be of considerable interest to examine the functional similarities and differences between these odors. The full ranges of these phenomena are not known and not understood.

The fact that some of these demonstrations have arisen from accidental experimental findings in the laboratory is of far-reaching and immediate significance. The olfactory capability of laboratory animals cannot be ignored. Indeed, it appears that much of the world of rodents is viewed through their noses. That world appears to be rich and complex beyond our imagination.

REFERENCES

Amoore, J. E. Stereochemical specificities of human olfactory receptors. *Perfumery and Essential Oils Record*, 1952, **43**, 321–323.

Amsel, A., Hug, J. J., & Surridge, C. T. Subject-to-subject trial sequence, odor trails, and patterning at 24-h ITI. *Psychonomic Science*, 1969, **15**, 119–120.

Archer, J. The effect of strange male odor on aggressive behavior in male mice. *Journal of Mammalogy*, 1968, **49**, 572–575.

Baran, D., & Glickman, S. E. "Territorial marking" in the Mongolian gerbil: A study of sensory control and function. *Journal of Comparative and Physiolgical Psychology*, 1970, **71**, 237–245.

Barnett, S. A. *The rat: A study in behaviour.* Chicago: Aldine, 1963.

Barnett, S. A., & Spencer, M. M. Responses of wild rats to offensive smells and tates. *British Journal of Animal Behaviour*, 1953, **1**, 32–37.

Beach, F. A. Analysis of the stimuli adequate to elicit mating behavior in the sexually inexperienced male rat. *Journal of Comparative Psychology*, 1942, **33**, 163–207.

Beach, F. A., & Jaynes, J. Studies of maternal retrieving in rats: I. Recognition of young. *Journal of Mammalogy*, 1956, **37**, 177–180. (a)

Beach, F. A., & Jaynes, J. Studies on maternal retrieving in rats: II. Effects of practice and previous parturitions. *The American Naturalist*, 1956, **90**, 103–109. (b)

Beach, F. A., & Jaynes, J. Studies of maternal retrieving in rats: III. Sensory cues involved in the lactating female's response to her young. *Behaviour*, 1956, **10**, 104–125. (c)

Bermant, G., & Taylor, L. Interactive effects of experience and olfactory bulb lesions in male rat copulation. *Physiology and Behavior*, 1969, **4**, 13–17.

Brill, M. Parameters of odor-trail avoidance in the spontaneous alternation of the rat. Unpublished doctoral dissertation, University of Cincinnati, 1967.

Bruce, H. M., & Parkes, A. S. An olfactory block to implantation in mice. *Journal of Reproduction and Fertility*, 1961, **2**, 195–196.

Calhoun, J. B. *The ecology and sociology of the Norway rat.* Washington, D.C.: Government Printing Office, 1962. (a)

Calhoun, J. B. Population density and social pathology. *Scientific American*, 1962. (b)

Carr, W. J., Krames, L., & Costanzo, D. J. Previous sexual experience and olfactory preference for novel versus original sex partners in rats. *Journal of Comparative and Physiological Psychology*, 1970, **71**, 216–222.

Carr, W. J., Loeb, L. S., & Dissinger, M. L. Responses of rats to sex odors. *Journal of Comparative and Physiological Psychology*, 1965, **59**, 370–377.

Carr, W. J., Loeb, L. S., & Wylie, N. R. Responses to feminine odors in normal and castrated male rats. *Journal of Comparative and Physiological Psychology*, 1966, **62**, 336–338.

Carr, W. J., Martorano, R. D., & Krames, L. Responses of mice to odors associated with stress. *Journal of Comparative and Physiological Psychology*, 1970, **71**, 223–228.

Carr, W. J., Wylie, N. R., & Loeb, L. S. Responses of adult and immature rats to sex odors. *Journal of Comparative and Physiological Psychology*, 1970, **72**, 51–59.

Carter, C. S., & Marr, J. N. Olfactory imprinting and age variables in the guinea pig, *Cavia porcellus*. *Animal Behaviour*, 1970, **8**, 238–244.

Curti, M. W. Native fear responses of white rats in the presence of cats. *Psychological Monographs*, 1935, **46**(6, Whole No. 210).

Davis, S. F., & Ludvigson, H. W. The "depression effect" and the problem of odor control. *Psychonomic Science*, 1969, **14**, 93–94.

Dominic, C. J. Observations of the reproductive pheromones of mice. I. Source. *Journal of Reproduction and Fertility*, 1966, **11**, 407–414.

Douglas, R. J. Cues for spontaneous alternation. *Journal of Comparative and Physiological Psychology*, 1966, **62**, 171–183.

Eisenberg, J. F. A comparative study of sandbathing behaviour in Heteromyid rodents. *Behaviour*, 1963, **22**, 16–23.

Eisenberg, J. F., & Kleinman, D. F. Olfactory communication in mammals, *Annual Review of Ecology and Systematics*, 1972, **3**, In press.

Gleason, K. K., & Reyneirse, J. H. The behavioral significance of pheromones in vertebrates. *Psychological Bulletin*, 1969, **71**, 58–73.

Goff, W. R. Measurement of absolute olfactory sensitivity in rats. *American Journal of Psychology*, 1961, **74**, 384–393.

Griffith, C. R. The behavior of white rats in the presence of cats. *Psychobiology*, 1920, **2**, 19–28.

Heimer, L., & Larsson, K. Mating behavior of male rats after olfactory bulb lesions. *Physiology and Behavior*, 1967, **2**, 207–209.

Howard, W. E., Marsh, R. E., & Cole, R. E. Food detection by deer mice using olfactory rather than visual cues. *Animal Behaviour*, 1968, **16**, 13–17.

Howard, W. E., Palmateer, S. D., & Marsh, R. E. A body capacitor olfactometer for squirrels and rats. *Journal of Mammalogy*, 1969, **50**, 771–776.

Kamin, L. J. Retention of an incompletely learned avoidance response: Some further analyses. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 713–718.

King, M. G. Stimulus generalization of conditioned fear in rats over time: Olfactory cues and adrenal activity. *Journal of Comparative and Physiological Psychology*, 1969, **69**, 590–600.

Klein, D., & Brown, T. S. Exploratory behavior and spontaneous alternation in blind and anosmic rats. *Journal of Comparative and Physiological Psychology*, 1969, **68**, 107–110.

Krames, L., Carr, W. J., & Bergman, B. A pheromone associated with social dominance among male rats. *Psychonomic Science*, 1969, **16**, 11–12.

Lee, S., van der., & Boot, L. M. Spontaneous pseudopregnancy in mice. II. *Acta Physiologica et Pharmacologica Neerlandica*, 1956, **5**, 213–215.

Le Magnen, J. Les phenomens olfacto-sexuals chez le rat blanc. *Archieves des Sciences Physiologiques*, 1952, **6**, 295–331.

Le Magnen, J. Olfaction and nutrition. In L. Beidler (Ed.), *Handbook of sensory physiology.* Vol. 4. *Chemical senses.* Germany: Springer Verlag, in press.

Long, C. J., & Tapp, J. T. Reinforcing properties of odors for the albino rat. *Psychonomic Science*, 1967, **7**, 17–18.

Long, C. J., & Tapp, J. T. An apparatus for the assessement of the reinforcing properties of odors in small animals. *Journal of the Experimental Analysis of Behavior*, 1968, **11**, 49–51.

Long, C. J., & Stein, G. W. An analysis of the reinforcing properties of food odour. *Canadian Journal of Psychology/Review of Canadian Psychology,* 1969, **23**, 212–218.

Ludvigson, H. W., & Sytsma, D. The sweet smell of success: Apparent double alternation in the rat. *Psychonomic Science,* 1967, **9**, 283–284.

Mackintosh, J. H., & Grant, E. C. The effect of olfactory stimuli on the agonistic behaviour of laboratory mice. *Zeitschrift fur Tierpsychologie,* 1966, **23**, 584–587.

Marr, J. N., & Gardner, L. E. Early olfactory experience and later social behavior in the rat: Preference, sexual responsiveness, and care of young. *Journal of Genetic Psychology,* 1965, **107**, 167–174.

Marr, J. N., & Lilliston, L. G. Social attachment in rats by odor and age. *Behaviour,* 1969, **33**, 277–282.

McCall, R. B., Lester, M. L., & Dolan, C. G. Differential rearing and the exploration of stimuli in the open field. *Developmental Psychology,* 1969, **1**, 750–762.

McCall, R. B., Lester, M. L., & Corter, C. M. Caretaker effect in rats. *Developmental Psychology,* 1969, **1**, 771.

McHose, J. H., & Ludvigson, H. W. Differential conditioning with nondifferential reinforcement. *Psychonomic Science,* 1966, **6**, 485–486.

Miller, S. D., & Erickson, R. P. The odor of taste solutions. *Physiology and Behavior,* 1966, **1**, 145–146.

Millman, B. Odor preferences and aversions in the rat. Unpublished masters thesis, University of Calgary, Canada, 1968.

Morrison, R. R., & Ludvigson, H. W. Discrimination by rats of conspecific odors of reward and nonreward. *Science,* 1970, **167**, 904–905.

Moulton, D. G. Studies in olfactory acuity: V. The comparative olfactory sensitivity of pigmented and albino rats. *Animal Behaviour,* 1960, **8**, 129–133.

Müller-Velten, H. Uber den Angstgeruch bei der Mausmaus (*Mus Musculus* L). *Zeitschrift fur vergleichende Physiologie,* 1966, **52**, 401–429.

Murphy, M. R., & Schneider, G. E. Olfactory bulb removal eliminates mating behavior in the male golden hamster. *Science,* 1969, **167**, 302–304.

Pain, J. F., & Booth, D. A. Toxiphobia for odors. *Psychonomic Science,* 1968, **10**, 363–364.

Pfaff, D., & Pfaffmann, C. Behavioral and electrophysiological responses of male rats to female rat urine odors. In C. Pfaffmann (Ed.), *Olfaction and taste: Proceedings of the third international Symposium.* New York: Rockefeller University Press, 1969.

Pranzarone, G. F. Male rats preference for estrous odor, or pheromone, as a function of differential exposure to females during postweaning development. Unpublished masters thesis, George Peabody College, 1969.

Ralls, K. Mammalian scent marking. *Science,* 1971, **171**, 443–449.

Reiff, M. Untersuchungen uber naturliche und synthetische Geruchstouffe, die bei Ratten und Mausen eine stimulierende Wirkung auslosen. *Acta Tropica,* 1956, **13**, 289–318.

Ropartz, P. The relation between olfactory stimulation and aggressive behavior in mice. *Animal Behaviour,* 1968, **16**, 97–100.

Rosen, J. J. Spontaneous alternation behavior in the rat. Unpublished masters thesis, Clark University, 1968.

Scott, J. W., & Pfaff, D. W. Behavioral and electrophysiological responses of female mice to male urine odors. *Physiology and Behavior,* 1970, **5**, 407–411.

Schapiro, S., & Salas, M. Behavioral response of infant rats to maternal odor. *Physiology and Behavior,* 1970, **5**, 815–817.

Seago, J. D., Ludvigson, H. W., & Remley, N. R. Effects of anosmia on apparent double alternation in the rat. *Journal of Comparative and Physiological Psychology,* 1970, **71**, 435–442.

Southall, P. F. Odor cues in a maze discrimination task: A re-evaluation. Unpublished masters thesis, Memphis State University, 1969.

Southall, P. F., & Long, C. J. Odor cues in a maze discrimination. *Psychonomic Science,* 1969, **16**, 126–127.

Steinbrecher, W. Die Duftwahl von Wander- und Hausratten. *Zeitschrift fuer Angewandte Zoologie,* 1962, **49**, 301–349.

Stern, J. J. Responses of male rats to sex odors. *Physiology and Behavior,* 1970, **5**, 519–524.

Still, A., & MacMillan, A. Odor trail and spontaneous alternation. *Psychonomic Science,* 1969, **16**, 160–161.

Sveredenko, P. A. Raziskeevanee grizunamee korma v zemle uslovnie refleksi nek na nepeeshchevie zapaxhee [Finding of food in the earth by rodents and their conditioned reflexes to non-food odors]. *Zoologeecheskee Jzurnal,* 1954, **33**, 876–888.

Tapp, J. T. Activity, reactivity and the behavior directing properties of stimuli. In J. T. Tapp (Ed.), *Reinforcement and behavior.* New York: Academic Press, 1969.

Tapp, J. T., & Long, C. J. A comparison of the reinforcing properties of stimulus onset for several sense modalities. *Canadian Journal of Psychology,* 1968, **22**, 449–455.

Tapp, J. T., Plemons, J., & Spiegler, M. Odor preferences in men and rats, unpublished manuscript, 1967.

Thiessen, D. D., Blume, S. L., & Lindzey, G. A scent marking response associated with the ventral sebaceous gland of the Mongolian gerbil (*Meriones Unguiculatus*). *Animal Behaviour,* 1970, **18**, 26–30.

Valenta, J. G., & Rigby, M. K. Discrimination of the odor of stressed rats. *Science,* 1968, **161**, 599–601.

Wasserman, E. A., & Jensen, D. D. Olfactory stimuli and the "pseudo-extinction" effect. *Science,* 1969, **166**, 1307–1309.

Whitten, W. K. Pheromones and mammalian reproduction. In A. McLaren (Ed.), *Advances in reproductive physiology.* New York: Academic Press, 1966.

Whitten, W. K., Bronson, F. H., & Greenstein, J. A. Estrus-inducing pheromones of male mice: Transport by movement of air. *Science,* 1968, **161**, 584–585.

(Received June 28, 1971)

Psychological Bulletin
1973, Vol. 79, No. 1, 45–47

RELATION BETWEEN SIGNAL DETECTABILITY THEORY AND THE TRADITIONAL PROCEDURES FOR MEASURING THRESHOLDS:

AN ADDENDUM[1]

MICHEL TREISMAN[2]

University of Reading, England

M. Treisman and T. R. Watts argued in 1966 that signal detectability measures could be related to those given by the traditional psychophysical procedures, and they presented evidence supporting this. P. D. Isaac has since criticized their method of analysis. The data are reexamined in light of this criticism. It is found that the original conclusion is strengthened, and a further implication of the data is noted.

Only if the independent variable has been measured with little or no error are we justified in fitting a regression line to joint observations (x_i, y_i) of variables whose true values are X_i and Y_i. If error terms affect both x_i and y_i and we assume the underlying random variables to be related by $Y = \alpha_s + \beta_s X$, the relation between the observations is given by:

$$y_i = \alpha_s + \beta_s(x_i - \epsilon_{x_i}) + \epsilon_{y_i}. \qquad [1]$$

This, which is known as a structural relation, is the appropriate model to fit, but although estimates for the parameters have been known since 1879 (Madansky, 1959), the structural relation has been largely ignored by psychologists, partly for reasons of unfamiliarity. A valuable contribution to dispelling this reason is offered by Isaac (1970), who summarized Kendall and Stuart's (1961) account of the maximum likelihood solution, and illustrated it by showing that Treisman and Watts (1966), who fitted a regression to their data, should have determined more appropriately a structural relation. Since this second point inevitably must cast doubt on results that have been thought to establish a theoretical point of some interest (D'Amato, 1970), the implications of this criticism require careful assessment.

It is sometimes taken that signal detectability theory (SDT) procedures have superseded those of classical psychophysics. Treisman and Watts (1966) argued that the same statistical decision mechanism, operating on normal distributions of stimulus effects, could underlie both types of task; if so, exact quantitative relations should exist between measures given by classical and SDT procedures. Accordingly, they derived and tested a prediction relating the method of constant stimuli and the SDT rating method. Two parameters are usually derived from the latter, d' and σ_{SN}/σ_N, where S = signal and N = noise. They showed that $\sigma_{\Delta I}$, the standard deviation of the method of constant stimuli psychometric function, corresponds to σ_{SN}, and that a statistic derived from the method of constant stimuli, $d'_{\text{est}} = (I_{SN} - I_N)/\sigma_{\Delta I}$, is related to d' by:

$$d'_{\text{est}} = (\sigma_N/\sigma_{SN})d'. \qquad [2]$$

This predicts a slope somewhat less than or equal to 1, since usually $\sigma_{SN}/\sigma_N \geq 1$. If we define a new SDT measure, $d'' = (M_{SN} - M_N)/\sigma_{SN} = (\sigma_N/\sigma_{SN})d'$, we can write this prediction more simply as:

$$d'_{\text{est}} = d''. \qquad [3]$$

This form is preferable, since it predicts the same slope for each subject, whereas σ_{SN}/σ_N will vary among subjects.

Using an auditory intensity-discrimination task, Treisman and Watts determined both d'_{est} and d' for 26 experimental sessions, the method of constant stimuli and the rating method being interleaved. They calculated the regression as:

$$d'_{\text{est}} = .77 + .77d'. \qquad [4]$$

---

[1] This research was supported by the Medical Research Council, England.

[2] The author would like to thank R. D. Stern for his comments.

Requests for reprints should be sent to the author, who is now at the Department of Psychology, University of Oxford, Oxford, England.

The slope did not differ significantly from .67, the reciprocal of the mean σ_{SN}/σ_N, nor from 1, but was significantly greater than 0. This appeared to confirm the prediction in Equation 2.

Issac (1970) correctly noted that Equation 2 describes a structural relation, both variables being measured with error, and therefore fitting such a relation would be most appropriate. Nevertheless, two points may be made in favor of the procedure followed. First, to estimate β_s requires prior knowledge of $\sigma^2_{\epsilon_x}$, or $\sigma^2_{\epsilon_y}$, the error variances, or $\lambda = \sigma^2_{\epsilon_y}/\sigma^2_{\epsilon_x}$, "a situation not too likely to arise in the biological sciences [Sokal & Rohlf, 1969, p. 483]" and which did not appear to arise here. Isaac (1970) suggested that λ can be estimated from the results of four subjects who repeated sessions, but this is doubtful: Each shows a positive relation between d'_{est} and d' over sessions. If, then, we fall back on regression analysis, we may note that it appears to give a conservative test of the null hypothesis (taking this to be the zero slope implied by a random relation) since the regression slope $\hat{\beta}_r = s_{xy}/s^2_x$ sets a lower bound for $\hat{\beta}_s$, which may be estimated by $\hat{\beta}_s = s_{xy}/(s^2_x - \sigma^2_{\epsilon_x})$.

But though the regression slope may be of service in rejecting the null hypothesis, it provides limited evidence on the form of the linear interrelation between the underlying variables (Isaac, 1970); therefore, a new analysis of the data was devised that allows λ to be estimated. Each session was divided into equal halves, and d'_{est}, d', and d'' derived for each half. (One session for which this could not be done was dropped.) The intrasession variation was taken as error variance and was used to estimate $\sigma^2_{\epsilon_y}$, $\sigma^2_{\epsilon_x}$ and λ, and the structural relation then was determined separately for d' and for d''. The first gave:

$$d'_{\text{est}} = .92 + .78(d' - \epsilon_{d'}) + \epsilon_{d'_{\text{est}}}. \quad [5]$$

The slope is significantly different from 0 ($t = 4.47$, $p < .001$) but not from .67 or 1. It is very similar to that in Equation 4, which does not, of course, imply that β_r will always approximate β_s closely; this is due here to the relatively low d' error variance indicated by the estimate for $\lambda = \sigma^2_{\epsilon_{d'\text{est}}}/\sigma^2_{\epsilon_{d'}}$, which was 7.49. The formula used for t was:

$$t = \left\{ (n-2) \sin^2 2(\hat{\theta} - \theta) \left[\frac{\frac{1}{4}(s_x{}^2 - s_y{}^2)^2 + s^2{}_{xy}}{s_x{}^2 s_y{}^2 - s^2{}_{xy}} \right] \right\}^{\frac{1}{2}} \quad [6]$$

where $\hat{\theta} = \tan^{-1} \hat{\beta}$ and $\theta = \tan^{-1} \beta_0$. (There is an error in the corresponding equation (Equation 29.62) in Kendall and Stuart [1961], and in Isaac's transcription of it.) As λ was calculated from the data, we took $df = n - 3$. The analysis for d'' gave:

$$d'_{\text{est}} = .93 + .99(d'' - \epsilon_{d''}) + \epsilon_{d'_{\text{est}}} \quad [7]$$

with $\lambda = 5.87$. The slope differs significantly from 0 ($t = 5.02$, $p < .001$) but not from 1. Both slopes are in good agreement with the predictions, and they differ from one another in the expected direction.

It seems that the estimated structural relations fully support the original theory. However, two findings deserve further comment. First, the high λ values indicate that the error variance for d'_{est} is considerably greater than that for the SDT measures. Second, the additive constants in Equations 5 and 7 are unexpectedly large, indicating that d'_{est} tends to be higher than d' or d''.

A suggestion can be made that may account for both findings. In applying SDT it is common to ignore possible variation in setting the criterion since this extra variance, σ^2_c, is not normally separable from the noise variance, σ^2_N. Nevertheless, it is likely that some criterion variance is always present. In the rating method, subjects had to maintain three separate criteria, and we suggest that σ^2_c was present at a constant and effectively irreducible level throughout the session. This would add to the noise variance, thus lowering d' or d'', with little difference between the half-sessions. In the method of constant stimuli, with only one criterion to maintain, we suggest that subjects were able to reduce σ^2_c for periods of time during the session. Thus σ^2_N and σ^2_{SN} would tend to be less in the method of constant stimuli than in the rating method, resulting in relatively higher d'_{est} values. When d' (d'') is particularly low, this sometimes must be due to an especially large component of criterion variance, making possible a correspondingly large reduction in σ^2_c in the method of constant stimuli with an equivalent rise in d'_{est}. This would tend to produce a positive intercept in Equations 5 and 7.

Furthermore, if the reductions in σ^2_c in the method of constant stimuli occur intermittently during a session, this will cause the intrasession variability in d'_{est} to be greater than that for the SDT measures—as shown by the λ values, each of which considerably exceeds 1 with $p < .001$. Since falls in σ^2_c would be more likely early in a session, subjects relaxing the extra effort as they get bored or fatigued, we might also expect that d'_{est} ,but not the SDT measures, would decline during the session. Though not reaching significance, the evidence is in this direction. Means for the two half-sessions were, for d', 2.07, 2.00 ($t = 0.28$); for d'', 1.68, 1.52 ($t = .71$); and for d'_{est}, 2.82, 2.22 ($t = 1.50$; $p < .1$, onetailed).

REFERENCES

D'Amato, M. R. *Experimental psychology*. New York: McGraw-Hill, 1970.

Isaac, P. D. Linear regression, structural relations, and measurement error. *Psychological Bulletin*, 1970, **74**, 213–218.

Kendall, M. G., & Stuart, A. The *advanced theory of statistics*. Vol. 2. *Inference and relationship*. London: Griffin, 1961.

Madansky, A. The fitting of straight lines when both variables are subject to error. *American Statistical Association Journal*, 1959, **54**, 173–205.

Sokal, R. R., & Rohlf, F. J. *Biometry*. San Francisco: Freeman, 1969.

Treisman, M. & Watts, T. R. Relation between signal detectability theory and the traditional procedures for measuring sensory thresholds: Estimating *d′* from results given by the method of constant stimuli. *Psychological Bulletin*, 1966, **66**, 438–454.

(Received July 23, 1971)

Psychological Bulletin
1973, Vol. 79, No. 1, 48–58

IMPRINTING IN PRECOCIAL BIRDS:
INTERPRETATION, EVIDENCE, AND EVALUATION[1]

D. W. RAJECKI[2]

University of Michigan

The identification of the basis for the phenomenon known as imprinting in young precocial birds is undertaken from six viewpoints. These are reviewed and examined in light of the experimental evidence concerning three aspects of imprinting: an early, nondiscriminative period of sensitivity to a wide range of parental surrogates, the formation of filial ties with a limited set of those surrogates, and the expression of filial bonds in subsequent situations. The concept of arousal emerges as a salient factor in virtually all interpretations of imprinting, but there is presently a divergence of opinion concerning the role that arousal plays.

The experimental investigation of imprinting in precocial birds has revealed several related patterns or aspects of the development of filial bonds in the young. These include (*a*) the tendency of newly hatched, naive birds to indiscriminantly approach a parent or parental surrogate, hereafter referred to as the target (this tendency rapidly declines with age); (*b*) the tendency of birds to exhibit increasing filial behavior toward a target presented repeatedly in a single exposure session or over more than one exposure sessions; and (*c*) the tendency of birds to exhibit subsequent filial behavior toward a previously exposed target when presented simultaneously with a novel target. The reports of the discovery of these elements have been ably reviewed by Bateson (1966) and Sluckin (1965).

These aspects find their counterpart in Evans' (1970) portrayal of the early and successive stages in the development of mobility and recognition in Ring-billed gulls (*Larus delawarensis*). Evans based his report on a large-scale study of imprinting in the semiprecocial gulls that included naturalistic observation, field experiments, and laboratory experiments. Over the first few days after hatching, three developmental stages emerge in terms of motor ability: (*a*) incipient (chicks remaining in or on the edge of the nest through the first day), (*b*) restricted (the onset of excursions limited to the vicinity of the nest site through the second and third days), and (*c*) extended (permanent emigrations, typically led by the parents, from the nest site to other areas from the fourth day on). Table 1 reproduces Evans' developmental scheme. It can be seen that the temporally related entries in the table under the columns headed by "Recognition" and "Approach responses" are akin to the aspects

TABLE 1

TEMPORAL RELATIONSHIPS BETWEEN DEVELOPMENTAL STAGES OF MOBILITY AND PROGRESSIVE RESTRICTION IN RANGE OF STIMULI EFFECTIVE IN ELICITING APPROACH RESPONSES IN YOUNG RING-BILLED GULLS

Developmental stage	Stimulus–response correlates of developmental stages	
	Recognition	Approach responses
I: Incipient mobility	"Innate propensity" to respond to a wide range of stimuli	To a wide range of initially effective stimuli
II: Restricted mobility	Species recognition	To a particular class of familiar stimuli
III: Extended mobility	Individual recognition	To stimuli associated with particular familiar individuals

Note.—Revised version reprinted with permission from a monograph by R. M. Evans published in *Animal Behaviour Monographs*, 1970, Vol. 3. Copyrighted by *Animal Behaviour Monographs*, 1970.

---

[1] Preparation of this paper was supported by a predoctoral Ford Foundation Fellowship administered by the Rackham Graduate School of the University of Michigan and in part by The National Science Foundation Grant GS-3119 to Robert B. Zajonc. The author is indebted to Robert B. Zajonc for critical comments on the manuscript.

[2] Requests for reprints should be sent to: D. W. Rajecki, Institute for Social Research, Research Center for Group Dynamics, University of Michigan, Ann Arbor, Michigan 48106.

of imprinting in precocial birds in the current usage. The subsequent evaluations of the interpretations of imprinting will be undertaken with two standards in mind: first, how well the interpretation accounts for the related elements of the phenomenon and second, how well the evidence supports the interpretation.

CONTEXTUAL INTERPRETATION OF IMPRINTING

Bateson (1966) cast an especially critical eye on the proposition that imprinting is unique in that it differs in kind or even in degree from other types of learning. It should be noted that for Bateson the term imprinting has a rather restricted usage: It is the *learning* of the characteristics of the parent or surrogate. Concluding that there is little basis for the claim of uniqueness (see also Hinde, 1970), Bateson did recognize that the phenomenon typically subsumed under the rubric of imprinting does have a "flavour of its own which needs to be accounted for without waxing mystical [p. 204]." Bateson suggested that rather than making claims for the process it might be more fruitful to examine the unique situation in which imprinting occurs.

In the first place, the newly hatched animal has a history of extremely limited experience and usually encounters a single target in an otherwise impoverished setting. Hence the preferences acquired early in life are achieved rapidly and have stability because of the absence of other acquired or competing preferences. But *preference* implies that appetitive or motivational factors are operating. The basis for this motivation is attributed to systems that have been established and that are operating prior to imprinting (learning). Bateson simply named one possible system as "filial behavior." This filial behavior is manifest in the observation that ". . . rather than passively filtering and classifying their (sensory) input, the birds actively search for a certain class of stimuli. Having located a member of that class, they direct their responses toward it and then begin to learn its characteristics [p. 208]." In other words, ". . . before their preferences have been restricted by imprinting the birds search for a broad range of conspicuous stimuli. On this view the consummatory stimuli that terminate this activity initiate imprinting and may be regarded as reinforcing. Subsequently the imprinting process restricts the range of consummatory stimuli, with the result that the birds only search for the object whose characteristics they have learned [p. 209]."

Contextual Interpretation Examined

One question to be asked of the interpretation is whether this pattern of events, the learning of the parental characteristics, is reconcilable with an instrumental or classical conditioning paradigm? Clearly, it is not an instrumental paradigm because the response, approach and following, does not require shaping and is emitted prior to conditioning. Is it then a classical conditioning paradigm? An unconditioned response is indeed present, approach and following, and this can become the conditioned response. An unconditioned stimulus is also present, the conspicuous stimulus, for this stimulus can automatically elicit the response on first presentation and hence reinforce the conditioned response to the conditioned stimulus. But if the unconditioned stimulus is the conspicuous stimulus, what is the conditioned stimulus? The *inconspicuous* elements of the stimulus in question? It seems unlikely.

If it is difficult to assign imprinting to one or another category of learning, it should at least be possible to predict consequences for performance if the reinforcement is removed. According to Bateson, the stimulus that terminates searching is the reinforcing event in the "learning" of the self same stimulus. It would seem that the removal of this reinforcing event should result in extinction. But extinction of what? Of the response? No, the response persists; the animal continues to search. Extinction of the retention of the stimulus characteristics? No, the animal will prefer the familiar to the novel in a subsequent choice.

In fairness it should be pointed out that apparently some structural change underlies the acquisition of the characteristics of the target by the imprinted bird. Bateson's later work shows that protein and RNA synthesis in the chick brain is affected by an imprinting procedure (Bateson, Horn, & Rose, 1969; Horn, Horn, Bateson, & Rose, 1971; Rose, Bateson, Horn, & Horn, 1970). It must also be pointed out that it is difficult to reconcile such acquisition in an imprinting situation with the conventional use of the global terms of stimulus, response, and reinforcement.

ETHOLOGICAL INTERPRETATION OF IMPRINTING

Although agreement is not universal among them, it has been proposed by ethologists (cf. Hess, 1970; Thorpe, 1956) that imprinting is one of any number of behaviors that have a fundamentally genetic basis. Such behavioral patterns are characteristic not of the individual but of the species to which the animal belongs and the *origins* of the behaviors are not determined or modified by experience, although the *expression* of the behaviors may come under environmental control. These innate behaviors are comprised of reflexes, taxes, and fixed-action patterns.

What triggers these innate behaviors? Not all perceived stimuli elicit a particular response. Rather, there is a certain limited range of stimuli termed "sign stimuli" or "releasers." Corresponding to the releasers are innate releasing mechanisms which respond to the external stimulation by providing energy for the performance of (or unblocking) the appropriate fixed-action pattern.

Thus imprinting in an ethological view is a process that occurs in the course of the expression of fixed-action patterns (approach and following) as released by certain sign stimuli (targets) through the operation of innate releasing mechanisms.

Ethological Interpretation Examined

The ethologists' concepts of species-related reflexes, taxes, and fixed-action patterns are worthy of study and provide an understanding of some behaviors. However, with special regard to the notion of sign stimuli the research on imprinting proves problematic.

Presumably, only a certain range of stimuli are effective in activating an innate releasing mechanism. On the contrary, the imprinting literature contains evidence that an extremely large assortment of objects and events will evoke the response. These include, in the visual realm alone: humans, balloons, electric trains (Fabricius, 1951, 1955; Fabricius & Boyd, 1954), footballs (Ramsay, 1951), small cubes and cylinders (Jaynes, 1956, 1957), colored cardboard boxes (Salzen & Sluckin, 1959), light flicker (James, 1959, 1960), moving horizontal lines (Smith & Hoyes, 1961), rotating discs (Smith, 1960), animals of a different species (Baer & Gray, 1960), and stationary three-dimensional objects (Gray, 1960), among others.

It is possible to argue, one supposes, that all of the above stimuli somehow happen to fall within the restriction of a genetically determined range. This supposition is weakened considerably by a report by Melvin, Cloar, and Massingill (1967). They found that 7 of 18 Bobwhite quail chicks imprinted on a live sparrow hawk. In fact, one of the chicks literally attacked and seized by the predator subsequently showed a strong tendency to approach the hawk. Moreover, the *responses* of young birds may be more labile than the ethologists would lead us to believe. In a tantalizing report on a single subject, Klopfer (1956) tells of a domestic chick (*Gallus gallus*) that not only imprinted on its cagemate, a swan goose (*Anser cygnoides*) gosling, but went on to adopt certain feeding habits and vocalization patterns of its peer!

In sum, the examination of imprinting to a formidable array of stimuli has revealed little on which to choose among them as sign stimuli. This is not to say that all stimuli are equally effective (see Gottlieb, 1971). It is only to say that the phenomenon cannot be understood merely in terms of the nature of the stimuli that elicit it.

CONDITIONING INTERPRETATION OF IMPRINTING

Moltz (1960) observed that Peking ducklings, first introduced to an imprinting apparatus containing a moving target when 8–10 hours of age, showed no emotionality and attended to the target only when it passed close by. However, when the same animal was returned to the same situation the next day, it was quite likely to become fearful and to begin shortly to follow the moving target. Now, when in proximity to the target, the animal ceased making distress calls and startle responses. From these observations Moltz made the following deductions:

1. During the initial exposure, some fraction of the animal's low-anxiety state becomes conditioned to previously novel stimuli: the target.

2. At the beginning of the second exposure, the animal's emotionality and excitement arises from having been placed in an unfamiliar situation. This unfamiliarity is based in part on the animal's observations of certain characteristics of the runway not previously observed (which is a result of its increased motor ability) and in part by having become familiar with some other situation, presumably the rearing area.

3. Given that the target has been previously associated with a low-anxiety state, it can function as a reinforcer on the second occasion and mediate new learning. A response instrumental in bringing the animal closer to the target should result in anxiety reduction, hence the animal should learn the response of approaching the target and remaining near it as long as the anxiety-producing situation continues to exist and the target continues to function as a reinforcer.

Conditioning Interpretation Examined

The conditioning interpretation holds that the attractiveness of any imprinting target, when the animal is in a state of stress, results from a *previous* association between the target and a certain visceral and cardiac state of the animal. This certain state is assumed to be a relatively positive one in that it is characterized by Moltz by the term "low anxiety." To the contrary, it has been found that relative to controls, birds show stronger attachment when *first* exposure to a target takes place under presumably noxious conditions. Pitz and Ross (1961) found that imprinting was enhanced when subjects were exposed to a loud noise when in a position to observe the target. Similarly, imprintable birds that received an optimal number of electric shocks on the initial trial showed stronger following responses when compared to nonshocked controls (Kovach & Hess, 1963). In a like finding, Salzen (1970) reported that chicks are particularly responsive to a target consisting of a flat board after they have been struck sharply by the board. Furthermore, Klopfer and Hailman (1964) reported that ducklings exposed to a 10-minute cold shock (40° Fahrenheit) immediately after initial exposure to an imprinting stimulus followed that stimulus 24 hours later as vigorously as nonchilled birds.

Although none of the studies cited report an objective measure of the noxious properties of the situations involved, similar levels are effective in avoidance conditioning in other species. These data, then, are not in support of the conditioning hypothesis advanced by Moltz (1960).

EPIGENETIC INTERPRETATION OF IMPRINTING

More recently, Moltz (1963, 1968) made the casual observation that it was quite likely that a Peking duckling would approach a moving target during initial exposure to it. Additionally, it seemed that the birds evidenced positive responses only when the target retreated from them, not when it approached. In order to account for this initial approach movement, Moltz appealed to an hypothesis advanced by Schneirla (1959) that in early stages in behavioral ontogeny, stimulus intensity rather than stimulus quality is a critical factor. Schneirla (1959, 1965) argued that early in development certain low or decreasing levels of stimulation automatically elicited approach responses while intense levels elicited withdrawal responses.

Moltz applied this formulation to the behavior of his subjects and concluded that, in terms of the maturational level of the ducklings, the retreating target provided the appropriate level of stimulation to elicit approach responses. The animals then followed the target as it continued to retreat in order to optimize this level of stimulation. A complementary observation was that if the target had occasion to move toward the subject, it did not elicit approach and sometimes elicited withdrawal responses.

Why, then, should the same target continue to elicit following on the second day? Presumably, the maturational state of the animal would have been radically altered while the stimulus value of the target remained constant. Moltz answered this question by again invoking the possibility that on the first exposure the *qualitative* characteristics of the target may very well become conditioned to the visceral states associated with approach processes, and attachment on subsequent tests is maintained by selective learning with fear reduction serving as the reinforcement.

Kovach (1970) considered the epigenetic position in a somewhat different light. He contrasted two possible explanations for the apparent increase, then decrease, in responsiveness of chicks to a novel target with increasing age: (*a*) The critical period analysis that holds that the maturational emergence of processes that support or compete with the approach response account for the behavioral disparities over age; (*b*) the epigenetic analysis that holds that these differences are functionally related to the animal's optimal arousal in terms of the animal's own sensitivity and the actual energy changes in the environment. In brief, early in life relatively high stimulus intensity should evoke approach, while later, lesser levels of intensity will be sufficient to evoke the same behavior. Thus if a target of constant intensity is presented to chicks of different ages, it should elicit most approach from the birds for which it produced an optimal level of arousal.

Epigenetic Interpretation Examined

This interpretation maintains that different quantitative levels of stimulation are optimal in eliciting approach behavior at various stages of the bird's maturation. Kovach (1970) put this proposition to a direct test by examining the influence on approach behavior of a number of factors including age, quantitative variation of the rate, intensity, and complexity of certain visual stimuli, and strain differences. Kovach found, for example, that the plot of the initial approach tendency of chicks of various ages (from 12 hours to 7 days) to a light flickering at different rates (from .8 to 8.9 cycles per second) took the form of an inverted U. That is, the intermediate frequencies were generally the most effective regardless of the age of the animal. Further, when responsiveness was plotted in terms of age with all frequencies combined, the inverted U was again observed. This relationship was also observed as a function of variations in luminance and complexity. Kovach concluded that the results argued more strongly for a critical period rather than an epigenetic interpretation of early approach behavior. He added, however, that some of the evidence could be best understood through an epigenetic analysis and that it was probable that elements represented by both formulations were influencing the animals' behavior.

Further, support is lacking for Moltz' (1963) position that retreating targets alone provide the amount of stimulation necessary to elicit approach responses. Moltz himself systematically tested the effect of the movement of the target and reported that while a retreating target was most effective, a target that only approached, or one that alternately approach and retreated, could reliably elicit approach responses on the part of the subject (Moltz, 1963). Salzen (1970) and Schulman, Hale, and Graves (1970) have reported like findings.

INFORMATION-PROCESSING INTERPRETATION OF IMPRINTING

Matthews and Hemmings (1963) undertook to provide an analysis of four "issues" emerging from the research on imprinting: (*a*) Why do chicks rapidly select particular objects in the environment? (*b*) Why do they follow these? (*c*) Why does this following last for only a limited period? (*d*) Why is there a correlation between following behavior and low anxiety cheeps and the cessation of following and overt signs of increased anxiety?

With regard to Issue *a* they begin with the premise that novelty is one of the characteristics that draws an organism's attention to a stimulus. Presuming that all stimuli are novel to newly hatched birds, they argue that some stimuli are more novel than others in terms of the number of variable dimensions they possess. Therefore, the greater the complexity, the greater the attraction. In lieu of the term complexity, however, Matthews and Hemmings chose to label this property the "information content" of the stimulation.

The basis for this attraction lies in the potential for the arousal of the bird. Assuming that the emergent bird is in a state of low arousal and low body metabolism, the animal's perception of the imprinting target would raise the arousal to a more satisfactory point and would increase ambulatory activity. Continuing to Issue *b*, should the target retreat or disappear, arousal level would be reduced to a less satisfactory point, and the animal would follow or search for the target to maximize or reestablish the input.

The following of a particular target would therefore cease (Issue *c*) when it had lost its capacity to sustain a satisfactory level of arousal, in other words, when it had become less novel. This could come about, according to Matthews and Hemmings, if one considers the receptor system as a limited capacity communications channel. A consequence of this limitation would be a recoding of the incoming stimulation (or information) in order to reduce the load on the channel. The upshot is that "the efficiency of this stimulation from the input begins to be reduced as the peripheral recoding process continues and a reduction in the efficiency of the reinforcer should be caused by the recoding process [p. 1184]." Thus the animal follows the familiar target less as it begins to seek out more novel stimulation. As for Issue *d* the animal gives evidence of anxiety when the target or the environment fails to provide adequate levels of stimulation.

Information-Processing Interpretation Examined

It is postulated that the attraction of an imprinting target for a subject will trace a certain temporal pattern. The target will initially evoke approach and following; after a time, however, this attraction will diminish and the subject will search for something that the target lacks.

This pattern was exemplified in the findings of a study by Moltz and Rosenblum (1958a). Peking ducklings were exposed to a moving target for 25 minutes on each of 15 consecutive days. For the first 5 days the subjects showed striking increases in the amount of time spent following the target. Their responses over the remaining days, however, showed an almost equal decline. Now, appealing to their hypothesis, Matthews and Hemmings (1963) would claim that up to Day 5 the target had provided sufficient novelty (information) to evoke satisfactory levels of arousal and hence following, but that after Day 5 the target had begun to lose this capacity and the ducklings followed it less. It remains, though, to examine a larger body of evidence before concluding that a target is effective only when *it* is the source of arousal.

For example, Moltz and Rosenblum (1958b) reported a study similar to the one in which they found a decline in following over later training days. In the more recent work they applied electric shock to some ducklings prior to the last 4 of 10 sessions. These showed superior following relative to nonshocked controls. Also, ducklings given an opportunity to adapt to the runway showed *less* responsiveness to the target compared to ducklings that were not adapted (Moltz, Rosenblum, & Halikas, 1959), and ducklings shocked in the runway (in the absence of the target) followed more than those shocked outside the runway (Moltz, 1963). Corollary findings are that the introduction of a strange object into the test area will restore the following behavior of birds that have ceased to follow the imprinting target (Candland, Nagy, & Conklyn, 1963; Sluckin & Salzen, 1961) and that the strength of following on second exposure is positively related to the length of time elapsed since the first exposure (Sluckin & Taylor, 1964).

In general, these findings indicate that arousal produced independently of the target enhances responsiveness toward a familiar object. What of the role of arousal in naive birds in the presence of a novel target? Adrenergic stimulants such as ephedrine and epinephrine (Kovach, 1964) as well as amphetamines (Kovach, 1964; Rajecki & Saegert, 1971) have been found to increase the likelihood and the strength of imprinting. Complementary results accrue from research wherein arousal is reduced via drugs. Hess (1959b) found that tranquilizers such as meprobamate and sodium pentobarbital reduce the incidence of approach behavior. Similarly, Bradford and Macdonald (1969) demonstrated that sodium pentobarbital, administered prior to training sessions, interfered with improvement of imprinting over trials in proportion to the strength of the dosage. It is important to report here that in subsequent tests both Kovach (1964) and Bradford and Macdonald (1969) discovered that a given degree of imprinting was not drug-state dependent; that is, differences observed under different drug conditions persisted in the absence of the drug at a later time.

It seems clear that arousal is a major influence on the elicitation and persistence of this social behavior. In light of the evidence presented, however, it is not at all clear that the target is the source of arousal as in the view of Matthews and Hemmings (1963). Although the evidence remains indirect, some writers (Hoffman, Stratton, Newby, & Barrett, 1970) maintain that even the initial presentation of an imprinting object has as a consequence the reduction of arousal as indicated by the cessation of distress calls in ducklings.

NEURONAL MODEL INTERPRETATION OF IMPRINTING

Imprinting for Salzen (1962, 1970) is the fixation of a class of objects to which social responses, both present and future, are directed and confined. Salzen accounted for the establishment of this fixation by applying Sokolov's (1960) neuronal modeling hypothesis for the orienting reflex in mammals. The hypothesis postulates a model-forming system in which neuronal assemblies, at the cortical level, are set up to correspond to incoming stimulation. If a subsequent incoming stimulation does not match the original model, the system excites an amplifying component, at the reticular level, which instigates the orienting reflex. Should the new afferent stimulation persist or be repetitive, however, a new neuronal model is established and this new model exerts an inhibitory influence on the amplifying component. The application of this hypothesis to imprinting is relatively straightforward, except that for Salzen orienting is made to include orienting toward the familiar and away from the novel.

In order to account for initial approach tendencies, Salzen argued that a characteristic of the newly hatched chick is that it has had extremely limited visual and auditory experience even though those systems are quite functional. This suggests to Salzen that any source of stimulation that meets a few simple requirements of the visual–auditory systems will instigate orientation and fixation or that the distance-receptors neuronal model systems are sufficiently unstructured so that any focal source of stimulation will give only a mild degree of mismatch and thus give mild activation in the form of orientation, fixation, and approach.

Then, as the length of exposure to the stimulation (target) increases, a corresponding refinement of the model results, and thus, repeated exposures lead to stronger indices of orientation (approach and following). However, a chick first exposed to a target at a relatively advanced age will initially find it discrepant with whatever distance-receptor models (perhaps of the rearing environment) that exist and therefore orient away from the unfamiliar. Thus the formation of neuronal models governs both the formation of attachments and the development of the avoidance of the novel.

Neuronal Model Interpretation Examined

A closer examination of the formulation of the neuronal model hypothesis (Sokolov, 1960) and the evidence that bears on it (reviews by Lynn, 1966; van Olst, 1970) reveals that Salzen's application is not entirely satisfactory. In the first place, the strength of Sokolov's hypothesis lies in its explanation not of the attractiveness of familiar stimuli, but of the *habituation* of the orienting reflex with repeated exposures of a stimulus. Second, habituation of the orienting reflex to a stimulus is thought to ultimately result in allowing an organism to ignore the stimulus and continued presentations may induce sleep (see Lynn, 1966). Third, the orienting reflex directs the organism's attention toward novel stimuli, not away. Finally, orienting is not necessarily correlated with approach; as Dimond (1970) points out, an organism may subsequently flee after orienting toward a stimulus.

It would seem, then, that the interpretation of imprinting in terms of the formation of neuronal models requires an alternative set of assumptions, possibly of the following sort: (*a*) At hatching there are absent or incomplete models for incoming stimulation with the consequence that the inhibition of the orienting reflex to random stimuli is not possible. (*b*) The persistence of orienting reflex to the general environment results in an arousal state noxious to the animal. (*c*) The presence of a salient stimulus (target) results in a rapid habituation of the orienting reflex to that stimulus. (*d*) The animal remains in proximity (approach and following) to the target to the degree that the state of habituation is preferable to the state of arousal elicited by the unfamiliar elements in the environment.

The best evidence for the neuronal modeling interpretation of imprinting, whatever form it takes, would be some direct measure of arousal. To date there is evidence not inconsistent with the hypothesis, but this evidence remains indirect. In Salzen's terms, the absence of highly structured models for the distance receptors in the just hatched, one should find very little "orienting away from" or fear of any stimulus configuration. Salzen (1970) traced the development of fear as a function of posthatch visual experience in several independent groups of chicks. Salzen's subjects were hatched in the dark and transferred to illuminated isolation cages when 1–12 hours of age. They were then subjected to an abruptly looming object (the experimenter's hand) immediately or at some later time (from 5, 15, or 60 minutes up to as much as 1 or 7 days) after the transfer. Salzen found virtually no response at all in the naive birds but as he tested over progressively greater lengths of visual experience, chicks began to flinch, then flee the object. Salzen concluded that the chicks with more experience were more disturbed because they had had a greater opportunity to form a neuronal model of the housing environment. It is interesting to recount here that Macdonald (1968) maintained chicks on a schedule of sodium pentobarbital injections for the first days after hatching. When tested on Day 4 and 5 these drug-treated birds showed a degree of imprintability far greater than birds similarly treated with saline. Macdonald (1968) concluded that the tranquilized birds remained sensitive to the imprinting target because the drug prevented them from imprinting on the environment.

An implication to be drawn from the neuronal model interpretation stems from the possibility that early posthatch distress (e.g., Collias, 1952) is due to the mismatch between contemporary stimulation and the neuronal models engendered by the pre- and perinatal environment. Presumably, if some element in the postnatal environment was like that of the prenatal, the animal would show positive responses toward it.

Some work on the specific influence of prenatal stimulation on postnatal behavior has proved inconclusive (Hess, 1959a; Oppenheim, 1968; Ramsay & Hess, 1954). Other investigators, however, have offered evidence that there is a direct relationship between prenatal experience and postnatal behavior. Gottlieb (1966) found that stimulation of neonate ducklings that involves the perinatal call does not have the effect of sensitizing the bird to maternal duck calls in general. Rather, it directs the bird's attention to the calls of siblings in the absence of conspecific maternal calls.

In a study employing contraspecific prenatal stimulation, Grier, Counter, and Shearer (1967) exposed chicken eggs to 1-second beeps of a 200-Hertz tone throughout Day 12–18 of incubation. A second batch of eggs was incubated in relative quiet. Within 6 hours of hatching all birds were tested for preference for sources emitting 200- or 2,000-Hertz tones. Grier et al. found that the chicks that had been exposed previously to the 200-Hertz tone chose to approach it and not the novel sound while the inexperienced chicks were indifferent in choice of approach to one or the other speaker. In a similar study dealing with a different sensory mode, Simner[3] established that in naive chicks there were preferred and nonpreferred rates of intermittent visual stimulation. However, when chicks received perinatal exposure to the nonpreferred rate, Simner found positively modified approach-eliciting properties for that rate.

---

[3] M. L. Simner, Effects of prehatch exposure to intermittent visual stimulation on posthatch preferences for intermittent visual stimulation in chicks. Paper presented at the Eastern Psychological Association Convention, New York, April 1971.

Dimond (1968) conducted a study that bears on the neuronal model interpretation. Half of a batch of incubating eggs were continuously exposed to a 60-watt light until just prior to hatching. The remaining eggs were incubated in darkness. All subjects were then hatched and reared in darkness until some time prior to being tested for responsiveness to a moving red balloon at 36 hours of age. One-third of the chicks were exposed to 15 minutes of posthatch illumination by a 60-watt bulb immediately before the imprinting test. Another third of the subjects received 6 hours of such illumination, and a final third received 12 hours of illumination. For the chicks receiving the shortest posthatch illumination, the dark-incubated subjects showed greater imprinting relative to the light-incubated group in terms of latency of first movement and time spent in proximity to the target. It may be possible to argue here that, at least in the shortest term posthatch illumination condition, the chicks in the group receiving prenatal illumination found the visual stimuli in the test situation discrepant with existing neuronal models for visual input, and thus initially avoided the target. This possibility is supported by the finding that for the balance of Dimond's subjects (the 6- and 12-hour posthatch treatment groups) absolute responsiveness and the difference in imprintability between the experimental and control groups (light- and dark-incubated) diminished as a function of increasing periods of illumination prior to testing.[4]

In sum, Salzen's neuronal model interpretation is plausible but requires amendment to adequately fit the data from imprinting research and the direct investigation of the orienting reflex in other species. The critical assessment of the position rests with future work.

CONCLUSION

Of the several interpretations considered here that do not merely dismiss the notion of imprinting, several (the ethological, conditioning, and epigenetic positions) have not received substantial support while at least one (the information-processing position) appears counter to some of the literature. It would appear that the neuronal model interpretation of imprinting is at least plausible when examined in the context of the aspects of imprinting. The idea that the avian foetus is actively influenced by pre- and perinatal events is long standing (Kuo, 1932a, 1932b, 1932c, 1932d). Further, the neuronal model interpretation has provisions for the explanation of the following several phases of precocial avian social behavior: the early sensitive period, the formation of social bonds, and the preference for the familiar over novel targets.

The clearest impression that emerges from literature, though, is the importance of arousal. In all six viewpoints noted, arousal is implicated in one or another sense. No doubt, writers have fallen back on this concept in the absence of another yeoman concept, reinforcement. However, arousal in an imprinting situation might be measured in terms of its physiological correlates. The curious thing is that, apparently, this has not yet been done. As seen, some would contend that imprinting involves an increase in arousal, while others make the same claim for decreased arousal. Before this issue can be decided, researchers on imprinting will have to examine the course of arousal throughout all aspects of imprinting.

REFERENCES

BAER, D. M., & GRAY, P. H. Imprinting to a different species without overt following. *Perceptual and Motor Skills*, 1960, **10**, 171–174.

BATESON, P. P. G. The characteristics and contexts of imprinting. *Biological Reviews*, 1966, **41**, 177–220.

BATESON, P. P. G., HORN, G., & ROSE, S. P. R. Effects of an imprinting procedure on regional incorporation of tritiated lysine into protein of chick brain. *Nature*, 1969, **223**, 534–535.

---

[4] Based on these findings it would appear that prenatal imprinting, at least prenatal auditory imprinting, is conceivable in a natural setting. This view is reinforced by the reports of Simms (1955) of a recording of a series of calls and responses between a brooding Stone curlew (a gregarious, essentially terrestrial shore bird) and her young still in the egg, and of Gottlieb (1963a) that a broody Wood duck hen begins to vocalize at a time that coincides with the pipping of her eggs. The onset of vocalizations occurs between 20–36 hours before the actual exodus of the clutch from the nest in the case of the duck.

Bradford, J. P., & Macdonald, G. E. Imprinting: Pre- and posttrial administration of pentobarbital and the approach response. *Journal of Comparative and Physiological Psychology,* 1969, **68**, 50–55.

Candland, D. K., Nagy, Z. M., & Conklyn, D. H. Emotional behavior in the domestic chicken (White Leghorn) as a function of age and developmental environment. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 1069–1073.

Collias, N. E. The development of social behavior in birds. *Auk,* 1952, **69**, 127–159.

Dimond, S. J. Effects of photic stimulation before hatching on the development of fear in chicks. *Journal of Comparative and Physiological Psychology,* 1968, **65**, 320–324.

Dimond, S. J. Visual experience and early social behaviour in chicks. In J. H. Crook (Ed.), *Social behaviour in birds and mammals.* New York: Academic Press, 1970.

Evans, R. M. Imprinting and mobility in young ring-billed gulls (*Larus delawarensis*). *Animal Behaviour Monographs,* 1970, **3**, 193–248.

Fabricius, E. Zur Ethologie junger Anatiden. *Acta Zoologica Fennica,* 1951, **68**, 1–175.

Fabricius, E. Experiments on the following response of Mallard ducklings. *British Journal of Animal Behaviour,* 1955, **3**, 122.

Fabricius, E., & Boyd, H. Experiments on the following reactions of ducklings. *Wildfowl Trust Annual Report, 1952–53,* 1954, **6**, 84–89.

Gottlieb, G. A naturalistic study of imprinting in Wood ducklings (*Aix sponsa*). *Journal of Comparative and Physiological Psychology,* 1963, **56**, 86–91.

Gottlieb, G. Species identification by avian neonates: Contributory effects of perinatal auditory stimulation. *Animal Behaviour,* 1966, **14**, 282–290.

Gottlieb, G. *Development of species identification in birds: An inquiry into the prenatal determinants of perception.* Chicago: University of Chicago Press, 1971.

Gray, P. H. Evidence that retinal flicker is not a necessary condition for imprinting. *Science,* 1960, **132**, 1834–1835.

Grier, J. B., Counter, S. A., & Shearer, W. M. Prenatal auditory imprinting in chickens. *Science,* 1967, **155**, 1692–1693.

Hess, E. H. Imprinting. *Science,* 1959, **130**, 133–140. (a)

Hess, E. H. The relationship between imprinting and motivation. *Nebraska Symposium on Motivation,* 1959, **7**, 44–77. (b)

Hess, E. H. Ethology and developmental psychology. In P. Mussen (Ed.), *Carmichael's manual of child psychology.* New York: Wiley, 1970.

Hinde, R. A. *Animal behaviour: A synthesis of ethology and comparative psychology.* New York: McGraw-Hill, 1970.

Hoffman, H. S., Stratton, J. W., Newby, V., & Barrett, J. E. Development of behavioral control by an imprinting stimulus. *Journal of Comparative and Physiological Psychology,* 1970, **71**, 229–236.

Horn, G., Horn, A. L. D., Bateson, P. P. G., & Rose, S. P. R. Effects of imprinting on uracil incorporation into brain RNA in the "split-brain" chick. *Nature,* 1971, **229**, 131–132.

James, H. Flicker: An unconditioned stimulus for imprinting. *Canadian Journal of Psychology,* 1959, **13**, 59–67.

James, H. Imprinting with visual flicker: Evidence for a critical period. *Canadian Journal of Psychology,* 1960, **14**, 13–20.

Jaynes, J. Imprinting: The interaction of learned and innate behavior. I. Development and generalization. *Journal of Comparative and Physiological Psychology,* 1956, **49**, 201–206.

Jaynes, J. Imprinting: The interaction of learned and innate behavior. II. The critical period. *Journal of Comparative and Physiological Psychology,* 1957, **50**, 6–10.

Klopfer, P. H. Goose-behavior by a White Leghorn chick. *Wilson Bulletin,* 1956, **68**, 68–69.

Klopfer, P. H., & Hailman, J. P. Basic parameters of following and imprinting in precocial birds. *Zeitschrift fur Tierpsychologie,* 1964, **21**, 755–762.

Kovach, J. K. Effects of autonomic drugs on imprinting. *Journal of Comparative and Physiological Psychology,* 1964, **57**, 183–187.

Kovach, J. K. Critical period or optimal arousal? Early approach behavior as a function of stimulus, age, and breed variables in chicks. *Developmental Psychology,* 1970, **3**, 73–77.

Kovach, J. K., & Hess, E. H. Imprinting: Effects of painful stimulation on the following response. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 461–464.

Kuo, Z. Y. Ontogeny of embryonic behavior in *Aves.* I. The chronology and general nature of the behavior of the chick embryo. *Journal of Experimental Zoology,* 1932, **61**, 395–430. (a)

Kuo, Z. Y. Ontogeny of embryonic behavior in *Aves.* II. The mechanical factors in the various stages leading to hatching. *Journal of Experimental Zoology,* 1932, **62**, 453–489. (b)

Kuo, Z. Y. Ontogeny of embryonic behavior in *Aves.* III. The structures and environmental factors in embryonic behavior. *Journal of Comparative Psychology,* 1932, **13**, 245–272. (c)

Kuo, Z. Y. Ontogeny of embryonic behavior in *Aves.* IV. The influence of embryonic movements on the behavior after hatching. *Journal of Comparative Psychology,* 1932, **14**, 109–122. (d)

Lynn, R. *Attention, arousal and the orientation reaction.* Oxford: Pergamon, 1966.

Macdonald, G. E. Imprinting: Drug-produced isolation and the sensitive period. *Nature,* 1968, **317**, 1158–1159.

Matthews, W. A., & Hemmings, G. A theory concerning imprinting. *Nature,* 1963, **198**, 1183–1184.

Melvin, K. B., Cloar, F. T., & Massingill, L. S. Imprinting of Bobwhite quail to a hawk. *Psychological Record,* 1967, **17**, 235–238.

Moltz, H. Imprinting: Empirical basis and theoretical significance. *Psychological Bulletin,* 1960, **57**, 291–314.

Moltz, H. Imprinting: An epigenetic approach. *Psychological Review*, 1963, **70**, 123–138.

Moltz, H. An epigenetic interpretation of the imprinting phenomenon. In G. Newton & S. Levine (Eds.), *Early experience and behavior*. Springfield: Charles C Thomas, 1968.

Moltz, H. & Rosenblum, L. A. Imprinting and associative learning: The stability of the following response in Peking ducks (*Anas platyrhynchos*). *Journal of Comparative and Physiological Psychology*, 1958, **51**, 580–583. (a)

Moltz, H., & Rosenblum, L. A. The relation between habituation and the stability of the following response. *Journal of Comparative and Physiological Psychology*, 1958, **51**, 658–661. (b)

Moltz, H., Rosenblum, L. A., & Halikas, N. Imprinting and level of anxiety. *Journal of Comparative and Physiological Psychology*, 1959, **52**, 240–244.

Olst, E. H. van *The orienting reflex*. The Hague: Mouton, 1971.

Oppenheim, R. W. Color preferences in the pecking response of newly hatched ducks (*Anas platyrhynchos*). *Journal of Comparative and Physiological Psychology Monograph Supplement*, 1968, **66**(3, Pt. 2).

Pitz, G. F., & Ross, R. B. Imprinting as a function of arousal. *Journal of Comparative and Physiological Psychology*, 1961, **54**, 602–604.

Rajecki, D. W., & Saegert, S. Effects of methamphetamine hydrochloride on imprinting in White Leghorn chicks. *Psychonomic Science*, 1971, **23**, 7–8.

Ramsay, A. O. Familial recognition in domestic birds. *Auk*, 1951, **68**, 1–16.

Ramsay, A. O., & Hess, E. H. A laboratory approach to the study of imprinting. *Wilson Bulletin*, 1954, **66**, 196–206.

Rose, S. P. R., Bateson, P. P. G., Horn, A. L. D., & Horn, G. Effects of an imprinting procedure on regional incorporation of tritiated uracil into chick brain RNA. *Nature*, 1970, **225**, 650–651.

Salzen, E. A. Imprinting and fear. *Symposia of the Zoological Society of London*, 1962, **8**, 199–217.

Salzen, E. A. Imprinting and environmental learning. In L. R. Aronson, E. Tobach, D. S. Lehrman, & J. S. Rosenblatt (Eds.), *Development and evolution of behavior*. San Francisco: Freeman, 1970.

Salzen, E. A., & Sluckin, W. The incidence of the following response and the duration of responsiveness in domestic fowl. *Animal Behaviour*, 1959, **7**, 172–179

Schneirla, T. C. An evolutionary and developmental theory of biphasic processes underlying approach and withdrawal. *Nebraska Symposia on Motivation*, 1959, **7**, 1–42.

Schneirla, T. C. Aspects of stimulation and organization in approach/withdrawal processes underlying vertebrate behavioral development. In D. S. Lehrman, R. A. Hinde, E. Shaw (Eds.), *Advances in the study of behavior*. New York: Academic Press, 1965.

Schulman, A. H., & Hale, E. B., & Graves, H. B. Visual stimulus characteristics for initial approach response in chicks (*Gallus domesticus*). *Animal Behaviour*, 1970, **18**, 461–466.

Simms, E. Conversational calls of birds as revealed by new methods of field recording. *Acta XI Congressus Internationalis Ornithologici*, 1955, 623–626.

Sluckin, W. *Imprinting and early learning*. Chicago: Aldine, 1965.

Slucken, W., & Salzen, E. A. Imprinting and perceptual learning. *Quarterly Journal of Experimental Psychology*, 1961, **13**, 65–77.

Sluckin, W., & Taylor, K. F. Imprinting and short-term retention. *British Journal of Psychology*, 1964, **55**, 181–187.

Smith, F. V. Towards a definition of the stimulus situation for the approach response in the domestic chick. *Animal Behaviour*, 1960, **8**, 197–200.

Smith, F. V., & Hoyes, P. A. Properties of the visual stimuli for the approach response in the domestic chick. *Animal Behaviour*, 1961, **9**, 159–166.

Sokolov, E. M. Neuronal models and the orienting reflex. In A. B. Brazier (Ed.), *The Central nervous system and behavior*. New York: Josiah Macy Foundation, 1960.

Thorpe, W. H. *Learning and instinct in animals*. Cambridge: Harvard University, 1956.

(Received July 26, 1971)

Psychological Bulletin
1973, Vol. 79, No. 1, 59–70

ATTENTION AND HEART RATE: A CRITICAL APPRAISAL OF THE HYPOTHESIS OF LACEY AND LACEY[1]

WILLIAM W. HAHN[2]

University of Denver

The hypothesis formulated by Lacey and Lacey concerning attention and heart rate is evaluated with respect to its original and subsequent expressions, its testability, and research results offering confirmation or disconfirmation of the hypothesis. Heart rate deceleration occurs in a variety of situations including many that require attention to the environment, but support for the instrumental role of this reaction in actively facilitating certain kinds of sensory-motor performance is lacking. The major contributions that are needed for the hypothesis to retain its heuristic value are clarification of constructs such as "attention" and greater explicitness of several aspects of the hypothesis, less reliance on heart rate alone as a dependent measure reflecting mediating mechanisms and cortical effects, and more investigation of other hypothesized mediating mechanisms.

Few theories have generated as much excitement in the field of psychophysiological research in the past 10 years as the Lacey formulation concerning the relationship between attentional variables and the heart rate response. Since its initial formulation, this theory has been the subject of considerable debate, and it has provided a framework for countless studies utilizing heart rate as a dependent variable. It may be useful at this time to attempt an appraisal of the Lacey position because some polarization has occurred in the field along an acceptance–rejection dimension, which may be reduced by a closer examination of the theory itself, as well as an appraisal of empirical results that are cited in support of the hypothesis.

WHAT DOES THE HYPOTHESIS SAY?

In its simplest and most frequently cited form, the hypothesis states that acceptance of the environment or attention to an external task is associated with a heart rate deceleration, whereas rejection of the environment, or attention to internal cues (associative processes) is associated with heart rate acceleration. However, the theory is really much more complex than this so that it must be examined in some detail before the various expressions of it can be appreciated.

One can observe the development of the Lacey hypothesis in their early concern with the general problem of measurement of autonomic responses and the relationships between characteristic autonomic patterns and motor impulsivity (Lacey, 1959; Lacey & Lacey, 1958). The neurophysiological relationships which later play a crucial role in the theory were first described as the basis for "autonomic-cortical coupling" in the Lacey and Lacey (1958) paper on motor impulsivity. At this time we see a determined attempt to link neurophysiological research results with behavioral phenomena while the authors courageously admit to trying to develop a theory.[3]

The first published presentation of the formalized hypothesis appeared in an article dealing with the application of psychophysiological measures to the study of psychotherapy (Lacey, 1959). In the process of describing how one autonomic indicator can show the impact of a stimulus while another does not, Lacey cited examples of heart rate or skin conductance failing to reflect the effects of a cold pressor test in some individuals. Hence, these measures yielded discordant answers to the question of the effectiveness of the stimulus in producing arousal. Subsequently, in a discussion of situational stereotypy Davis' findings (Davis, 1957; Davis & Buchwald, 1957; Davis, Buchwald, & Frankman, 1955) regarding unexpected changes in the direction of a response are cited and original data are presented on directional fractionation of heart rate and galvanic skin responses (GSR). After presenting these data, Lacey proceeded to offer some speculations about the reasons for the observed directional fractionation. A brief review of the blood pressure, carotid sinus, cortical-inhibiting mechanisms led to the first rudimentary expression of the hypothesis:

---

[1] The author wishes to thank Joseph Campos, Paul Obrist, John Stern, and Jay Trowill for their helpful suggestions and encouragement in the development of this paper.

[2] Requests for reprints should be sent to William W. Hahn, Department of Psychology, University of Denver, University Park, Denver, Colorado 80210.

[3] At various times these formulations are referred to as theory or hypothesis or hypotheses, and it is seldom clear where theory ends and hypotheses begin. We have not attempted to clarify these distinctions because of our reluctance to make differentiations that the authors themselves have not made.

> . . . where cardiac deceleration is the rule, the subject is required primarily to note and detect the environment. . . . At the other end, where cardiac acceleration is the rule, one can at least speculate that the opposite of environmental intake is called for. In a sense, then, the acceleration or deceleration of the heart could be considered to be something like an instrumental act of the organism, leading either to increased ease of environmental intake or to a form of "rejection of the environment" [Lacey, 1959, p. 199].

Four years later the hypothesis began to lose the status of "speculative notions" and to take on added stature as a major hypothesis:

> It is not too great an extrapolation to hypothesize that "mental concentration" is accompanied by cardiac acceleration, and that attention to the environment is accompanied by cardiac deceleration [Lacey, Kagan, Lacey, & Moss, 1963, p. 174].

In the same paper the hypothesis is developed further:

> We interpreted the results to mean that high heart rates were inhibitory and low heart rates facilitatory, for sensory-motor impulsivity. . . . [and] We interpreted these results to mean that an increase in [skin] conductance was excitatory and could overcome the presumably inhibitory effects of a high heart rate, but had no detectable effect when heart rates were low, this latter state itself being excitatory [p. 176].

A final expression of the hypothesis offers a rewording of it which adds at least one new dimension and has produced some confusion in attempts to test the hypothesis. Again, in Lacey et al.'s own words,

> The first study tested a hypothesis . . . that heart rate will decrease . . . when the subject sustains attention to the external environment . . . when the subject *"wants to accept"* environmental inputs. . . . The hypothesis also required acceleration be observed when the subject *"wants to reject"* environmental input . . . [pp. 188–189, italics added].

This hypothesis obviously offers a number of difficulties for empirical testing. It would seem to require manipulation and measurement of various combinations of acceptance and attention—difficult cognitive concepts which should be explored separately and in combination but probably cannot be. How can one pay attention and neither accept nor reject? Is inattention the same, or different from, rejection of the environment? And so on. When other variables, such as the "override" by GSR and the volitional or motivational variable of "want" to reject, are added, the hypothesis is of dubious testability by current methods. Yet, the hypothesis has received a great deal of testing and a large amount of evidence has been presented in *confirmation* of the hypothesis! How can this occur? For one thing, investigators have "zeroed in" on one or two aspects of the theory, namely, the acceptance–rejection dimension and heart rate change; while other factors, such as the role of GSR, have been largely ignored. Also, it is a flexible hypothesis which can accommodate the results of a variety of experiments, and finally, certain aspects of the hypothesis were based on reliable and replicable research results which numerous investigators have continued to report.

Later explications of the theory have not simplified or clarified these issues but rather have broadened its scope. For example, it is suggested that the cardiac deceleration leads to *sensory* facilitation; barorecptor cardiac changes produced within a single cardiac cycle also produce cortical arousal or inhibition. Not only carotid sinus baroreceptors but also aortic baroreceptors and possibly others also contribute to the mediation process and produce cortical facilitation or inhibition (Lacey, 1967).

In the most recent offering of both theory and data, the heart rate deceleratory response is found to be positively correlated with the "contingent negative variation" response which Lacey and Lacey (1970), citing the work of W. Grey Walter and his colleagues, characterize as a measure of an organism's "readiness to respond." While it is important to relate the heart rate mechanisms to other physiological responses of behavioral significance, it is also regrettable that this latest version of the hypothesis introduces new issues such as vicarious stress and reintroduces old issues such as response intention and the relationship of cognitive style to cardiac deceleration. At times, one has the feeling that the hypothesis grows too quickly and always outward—expanding to encompass new concepts and measures before results from earlier investigations can be used to help settle some of the original issues.[4]

NEUROPHYSIOLOGICAL EVIDENCE

In 1958 Lacey described neurophysiological studies which delineated the pathways from carotid sinus stretch receptors to brainstem mechanisms. The effects of these baroceptor mechanisms on cortical electroencephalogram (EEG) were also outlined. This work provided the neurophysiological foundation for the theory proposed in 1959 in that it demonstrated that blood pressure changes mediated by baroreceptors in the aortic arch and carotid sinus produce alterations in reticular formation activity and in cortical excitability. It had been shown that artificial distension of the carotid sinus of the Flaxedilized cat changed an activated EEG pattern into one of drowsiness or sleep (Bonvallet, Dell, & Hiebel, 1954). At about the same time it was shown that the slow wave patterns produced by adrenaline or noradrenaline and the cortical excitation produced by hypotensive drugs were abolished after sinoaortic denervation (Nakao, Ballim, & Gellhorn, 1956). Dell (1957) and Dell, Bonvallet, and Hugelin (1954) suggested that the stretch receptors of the carotid sinus produce an inhibitory effect on central vasomotor tone and adrenal secretion by way of Hering's nerve, and demonstrated that sectioning of this nerve abolished the inhibitory effect produced by an injection of epinephrine. By demonstrating this effect in a prebulbar sectioned preparation, they concluded that the inhibition occurred in the bulbar inhibitory reticular formation. Bonvallet and Allen (1963) provided the final link in this chain of events by showing that interruption of cardiovascular afferent input (by sectioning of the glossopharyngeal and vagus nerves) to *n. tractus solitarius* (nTS) led to a prolonged EEG activation pattern following reticular formation stimulation. This suggested that the normal inhibitory action of the carotid sinus reflex was lost due to interruption of the feedback information on which this inhibitory influence normally depends.

At this point it would seem reasonable to conclude that both tonic and phasic changes in *blood pressure* or pulse pressure, acting on carotid sinus baroreceptors, can produce observable changes in reticular formation and in the cortical electroencephalogram. However, it should be noted that these results do not bear very directly on the principal concerns of Laceys' developing hypothesis in which the acceptance–rejection task dimension and changes in *heart rate level* are the variables explored. Neurophysiological corroboration of Lacey's hypothesis would involve a study which manipulates acceptance–rejection in human subjects, while obtaining measures of skin conductance, heart rate, blood pressure, carotid sinus and aortic baroreceptor output, inhibitory influence in Hering's nerve, activity at nTS, and the cortical EEG. This experiment, for obvious reasons, has not yet been performed. The need for chronically implanted recording electrodes may require the use of infrahuman subjects, and this would necessitate considerable sacrifice in ability to equate an acceptance–rejection dimension with that stimulus-organism characteristic in the human studies.

---

[4] For example, a recent review article by Elliott (1972) cites a number of attempts to test the hypothesis that have produced negative results. Some of the studies are carefully constructed and the negative results probably cannot be dismissed on the basis of methodological or technical weaknesses. It would be interesting to see the Laceys' analysis of these studies and, if necessary, modifications of their theory at points where relevant data does not offer confirmation of it.

Psychophysiological Evidence

There are several principle lines of development of research relating to Lacey's hypothesis. The Laceys, turned to studies of reaction time and heart rate deceleration during the foreperiod, while others began to investigate alternative explanations for the reported heart rate changes or added new dimensions, such as task difficulty and verbalization requirements, to the already complicated hypothesis.

In 1963 Obrist reported a replication of the major results previously reported by the Laceys. The intake-rejection variable was manipulated by use of different tasks, while skin conductance, heart rate, blood pressure, and respiration were measured. The two major hypotheses tested (that tasks requiring close attention would result in depression of cardiovascular activity and that a noxious and a conceptual task would result in sympatheticlike changes in the cardiovascular system) were supported by the results. The author was evidently impressed with this confirmation of the theory:

> The present experiment confirms the previous work by Lacey on cardiovascular differentiation of various stimuli and heart rate acceleration to others [Obrist, proposed theoretical model. The direction of the cardiovascular changes in response to any one stimulus among subjects is very consistent; 28 of 28 subjects demonstrate heart rate deceleration to certain stimuli and, in most respects, is consistent with the 1963, p. 456].

The next presentations of data came in the form of three papers delivered at professional meetings.[5] The Lacey and Lacey paper, "Cardiovascular and Respiratory Correlates of Reaction Time," reported three experiments which explored the relationships between heart rate variability, heart rate levels, and reaction time performance. Using a fixed foreperiod of 4 seconds, they found a significant decrease in heart rate just prior to stimulation in 74 of 78 subjects and a rise in heart rate following stimulus presentation which led to characterization of this response as a V-shaped response. Over the entire experiment there was a correlation of $-.27$ between heart rate deceleration and reaction time, a correlation which rose to $-.39$ if only trials on which the subject was inspiring at time of stimulation were evaluated. No significant correlation between heart rate level and reaction time was found.

To determine if individual subjects demonstrated a significant relationship between reaction time and heart rate deceleration the authors attempted to partial out the common variance due to time factors by using data only from the second half of the experiment and excluded all subjects who showed a significant correlation between trial number and either heart rate or reaction time. Of the 63 correlations computed, 50 were negative and 17 of these were significant at $p < .05$ or less. The authors state that the probability of obtaining this many significant correlations is less than .001. However, knowing that 15 of the original 78 subjects showed significant correlations between trial number and these variables, one wonders to what extent the remaining 63 correlations were influenced by this third factor of time or learning. That is, even though the correlations were insignificant the fact that heart rate and reaction time varied with trial number to some degree may have assisted "chance" in producing the 17 out of 63 significant negative correlations. Removing the subjects who showed *significant* correlations between trial number and reaction time or heart rate does not remove the influence of this variable from the correlations between reaction time and heart rate. The same argument might apply to the analysis used subsequently to demonstrate that heart rate level is significantly correlated with reaction time for individuals. Furthermore, it would not be unreasonable to suggest that heart rate might decrease during the course of the experiment due to habituation, while reaction time performance might be expected to improve due to practice. Such long-term changes could contribute to the observed relationships when data are sampled from individual trials over the test period. It would seem preferable to have looked at the heart rate deceleration and reaction time performance when trial number was partialed out using covariance techniques.

---

[5] These papers were: (*a*) B. Lacey and J. Lacey, Cardiovascular and respiratory correlates of reaction time. Paper presented at the annual meetings of the Society for Psychophysiological Research, Washington, D.C., October 1964. (*b*) B. Lacey and J. Lacey, Change in cardiac response and reaction time as a function of motivation. Paper presented at the meetings of the Society for Psychophysiological Research, Denver, October 1966. (*c*) J. Coquey and J. Lacey, The effect of foreperiod duration on the components of the cardiac response during the foreperiod of a reaction time period. Paper presented at the meetings of the Society for Psychophysiological Research, Denver, October 1966. An expanded version of the first paper was included in National Institute of Mental Health Progress Report MH 00623, June 1965, as appended Preprint No. 3. The second two papers are summarized briefly in Lacey and Lacey, 1970.

The only substantiated conclusion to be derived from the second study is that the subjects again showed cardiac deceleration during the foreperiod; 10 out of 11 subjects showed a significant deceleration. Of these 10, 5 also showed a significant blood pressure increase (median for the group = 1 millimeter Hg). No significant group differences were found between either blood pressure or heart rate and reaction time performance, while correlations for individuals showed a marked preponderance of negative correlations between reaction time and cardiac deceleration.

In the third study, a reaction time task with a 9-second foreperiod was used to guard against the possibility that the 4-second foreperiod of the previous studies was capitalizing on an inherent cardiac variability response of about 4 seconds in duration. The cardiac deceleration was again observed in a majority of subjects (34 of 42); no evaluation of relationships between heart rate and reaction time was presented.

From the Lacey and Lacey paper, we might conclude that there is a slight relationship, for a group of subjects, between heart rate deceleration and reaction time. This relationship can be expressed by correlation and varies from − .06 to − .42 if different task conditions and respiratory phase are computed separately; it averages out to − .27 over all conditions for one entire experiment. These results only suggest that averaged data reflect an association between heart rate deceleration during the foreperiod and faster reaction times; they do not necessarily support an instrumental role of the heart rate response.

The two papers presented in Denver in 1966 (see Footnote 5) showed that the heart rate deceleration obtained was present under different stimulus conditions and that it was significantly enhanced by conditions designed to produce more motivation in the subjects.

A number of investigators have examined the theory in a developmental context (e.g., Chase, Graham, & Graham, 1968; Lewis, Kagan, Campbell, & Kalafat, 1966; Steele & Koons, 1968; Steele & Lewis, 1968). A detailed analysis of these studies is beyond the scope of this review. However, it should be noted that investigators have reported heart rate decelerations occurring in a wide variety of tasks requiring the child's attention, ranging from simple stimuli eliciting orienting reactions to conditioning and problem-solving tasks. The results from this area of inquiry provide some of the strongest supporting evidence and enthusiasm for the Lacey hypothesis, although practically none of these investigations examine the instrumental role of the cardiac response nor do they measure the response of hypothesized mediating mechanisms.[6]

Obrist and his collaborators have published a series of studies on reaction time and heart rate. Due to results of earlier studies on the relationships between electromyogram (EMG) and heart rate during classical conditioning procedures, these investigators have evaluated EMG, heart rate, and respiration during a reaction time task with a fixed foreperiod (Obrist, Webb, & Sutterer, 1969; Webb & Obrist, 1970). A striking concomitance was found in each of the measures; normalized scores for EMG, respiration frequency, and heart rate showed an initial increase from base line with onset of the ready signal followed by a marked decline in all measures which continued up to the time the respond signal was given. Correlations between heart rate change and respiration amplitude, frequency and EMG varied from .25 to .39 and the correlation between mean reaction time and amount of heart rate deceleration was .31. The EMG change and the respiratory frequency change correlated with mean reaction time, .42 and .37, respectively. It was also reported that the chin EMG appeared to reflect a more generalized form of inhibited motor activity, as it could be extended to other EMG measures, and eye blinks and eye movements also decreased in frequency along with heart rate and EMG during the foreperiod.

---

[6] For a more thorough treatment of this subject matter see the recent review by Graham and Jackson (in press).

Generally, Obrist and his collaborators have consistently reported a positive relationship between EMG measures and heart rate, for example, both decreasing during the foreperiod, and a negative correlation between these responses and reaction time, that is, reaction time performance improving or latency decreasing. The heart-rate–reaction-time relationship is, of course, consistent with the Lacey formulation, and the difference of opinion arises from different views about the functional or instrumental aspects of this relationship:

> Lacey (1967) has cited these studies as evidence for a feedback mechanism which related cardiac deceleration to cortical arousal via a reduction in baroreceptor activity. An alternative model . . . views the cardiac deceleration as part of a somatic inhibitory pattern [Webb & Obrist, 1970, p. 390].

In recent experiments these investigators have attempted to provide a critical test of the relaxation versus the inhibitory model by striving to block the deceleratory heart rate response. The reasoning seems to be that if the heart rate deceleration is not allowed to occur but the EMG decrease and enhanced reaction time performance are still obtained, the heart rate must not play a crucial role in instrumentally facilitating the reaction time performance. While earlier attempts did not supply the sought after differentiation "because the only combination of variables which eliminated HR deceleration also eliminated the inhibition of somatic measures [Webb & Obrist, 1970, p. 389]," a recent study utilizing pharmacological blocking of the heart rate response has produced the desired results (Obrist et al., 1970). This study demonstrated that atropine blocking of the heart rate deceleration did not reliably alter reaction time performance, and the authors concluded that heart rate is not significantly involved in these relationships within the context of the design of this particular kind of study. In this experiment, the authors have provided an important test of the general somatic inhibition versus feedback models. While minor methodological problems exist (e.g., it is not certain that initial assumptions regarding the effects of various preparatory intervals on heart rate must be accepted, nor is it clear that the effects of atropine are evaluated as completely as possible), the major results contribute to a serious weakening of the instrumental role of heart rate deceleration in facilitating reaction time performance. While these results weaken one major aspect of the Lacey hypothesis, they do not invalidate the entire hypothesis; as we have pointed out above, and as Obrist et al. (1970) rightly conclude, the afferent discharges from the aortic and carotid baroreceptors have not been measured. It is still possible that pulse-pressure changes were occurring at these sites independent of heart rate change, and these pressure changes may have produced the reaction time facilitation observed. Nevertheless, these studies do raise serious questions about the utility of using heart rate as the only indicator of the hypothesized facilitatory mechanisms and suggest that if these mechanisms play a role in such situations, they must do so independently of changes in heart rate.

The series of studies by Obrist and his collaborators have helped to clarify the issues and to correct one popular oversimplification of the Lacey hypothesis—that heart rate alone can be used to reflect the complex pressure changes at baroreceptor sites and hence the release of inhibition at nTS. In retrospect, it would seem that an investigator setting out to test the hypothesis would not have been content with less than a measurement of the afferent output over the ninth and tenth nerves and/or activation of the nTS mechanism; but the provocative early results with heart rate, the limitations of working with human subjects, and the difficulty of taking even pulse-pressure measurements seemed to encourage and proliferate the use of heart rate as a principal determinant of the hypothesized mediating mechanisms.

A second major group of studies has

focused on the role of verbalization, a variable not originally treated by the Laceys (at least not intentionally). These studies demonstrate the importance of the effects of instructions to verbalize on the heart rae response and point to the possibility of the confounding of verbalization with acceptance–rejection. (see Campos & Johnson, 1966; Campos & Johnson, 1967; Edwards & Alsip, 1969; Johnson & Campos, 1967). In two of the studies, repeated failures to find directional fractionation are not impressive because the stimuli used are different than those used by Lacey and often the lack of directional fractionation was based on a finding of no significant changes in either skin conductance or heart rate (in one study only 8 of 18 comparisons reached the .01 level of significance in either direction; in another, 6 out of 16), suggesting that the stimuli being used may not have produced *any* reliable effect in the responses measured.

Evidence from Studies on Related Issues

There are a number of studies, such as those involving classical conditioning of heart rate which have not been designed to test the Lacey hypothesis, but which do reflect on certain aspects of the hypothesis. For example, a finding of a decelerative anticipatory heart rate in conditioning studies is consistent with the Lacey formulation (if one assumes that the conditioned stimulus, CS, calls the subject's attention to the environment), and is not consistent with either a substitution view of classical conditioning or a view of activation theory which would propose that an acceleratory heart rate should be associated with the fear or arousal associated with the CS preceding a noxious unconditioned stimulus (UCS). Thus, the results of such studies may provide additional evidence on the various physiological reactions taking place during a foreperiod and, thereby, add information about the role of the mediating mechanisms proposed by the Laceys.

One example of this kind of investigation may be given. In a study of heart rate conditioning in the cat (Hein, 1969), an extensive battery of physiological measures was taken and each assessed during a CS+ and CS− in a classical conditioning format with shock used as the UCS. While marked variability in the several response measures occurred among the 20 experimental animals, all eventually achieved a high degree of stability of the decelerative-conditioned heart rate response as well as differentiation of the CS+ and CS−. Some of the results relevant to mechanisms proposed by the Laceys were: (*a*) Blood pressure decreased along with the heart rate decrease in response to the CS. (*b*) There were several signs of activation concomitant with the conditioned decelerative heart rate response including activated cortical EEG, pupillary dilatation, and galvanic skin potential. (*c*) Other concomitants of the CR were considered to be signs of parasympathetic activity or deactivation; these included heart rate deceleration itself, blocking of 40–50 cycles per second spindles in the olfactory bulb, and muscular relaxation. While Hein does not relate his results to the Laceys' hypothesis, it may have considerable import for this formulation. For example, here we see the often-neglected blood pressure measurement reflecting a fall during the foreperiod concomitant with cortical EEG activation just as Lacey has hypothesized,[7] while at the same time, the muscular relaxation observed consistently by Obrist and his collaborators and at least one other sign of deactivation are also reported. Hein's (1969) comment that ". . . it is difficult to conceive of a single 'center' that could trigger off such a complex response . . . [p. 462]" might be generalized to provide a cautionary note for the whole controversy over the Laceys' hypothesis. There is no reason why all of these responses cannot occur at one time, and there is no particular reason to focus on a single issue or a response measure or two, such as only the baroreceptor reflex mechanisms or primarily the heart-rate–EMG coupling phenomena. Aside from an artificially inflated valuing of parsimony, there appear to be no real advantages to conceptualizing the complex processes that occur in situations like these in such simplified terms; on the other hand, the use of multiple measures of central and peripheral physiological responses may provide a more accurate and more informative picture of the events taking place.

---

[7] Hein (1969) says, "The blood pressure does not rise during the deceleration and thus the carotid sinus reflex is not implicated [p. 461]." Evidently, what is meant here is that an initial rise in blood pressure did not cause the heart rate deceleration through the carotid sinus baroreceptor compensatory reflex. This would not, presumably, imply that the concomitant decreases in heart rate and blood pressure could not participate in the cortical activation through carotid sinus reflexes.

Another link in the chain of mediating mechanisms is developed by study of the relationship between discrete phases of the cardiac cycle and various measures of cortical arousal or performance. There have been reports of a significant relationship between a specific phase of the electrocardiogram (EKG) cycle and the number of erroneous psychomotor responses (Callaway & Yeager, 1960), auditory reaction times (Birren, Cardon, & Phillips, 1963), evoked responses (Callaway & Layne, 1964), and auditory thresholds (Saxon, 1970).

The complications introduced by intrinsic ties between preparatory interval and EKG cycle length has been the subject of intensive study (Thompson & Botwinick, 1970). In a series of four experiments, these investigators repeatedly find strikingly negative and insignificant results when the effects of the preparatory interval are controlled by various statistical or methodological techniques. The authors note that the discrepancy between their findings and earlier positive results may be due to the fact that a variable intertrial interval was used in their study or to the individual differences among subjects; but one must agree with Thompson and Botwinick that if the phenomenon is dependent on such fine adjustments of methodology, its utility and generality are of questionable value. Space limitations will not permit a more detailed analysis of this problem; however, it should be noticed that there is neurophysiological evidence which suggests that the number of cells in nTS firing in concert with the EKG rhythm and the reticular concomitants of these effects may be far too small to account for the presumed cortical activation. For the interested reader this theme is treated in more detail in a forthcoming paper.[8]

APPRAISAL

One of the major difficulties encountered in attempting to test or evaluate the hypothesis centers on the use of the intervening variable of "attention." Not only are there inherent definitional problems associated with this term, which is seldom adequately operationalized, but in addition, there is evidence that effects of attention may be confounded by verbalization and arousal factors and there is a marked temptation to reinterpret the unmeasured attentional factors after results become known. It is unfortunate that this key construct is also one of the least testable and most poorly measured aspects of the theory. Others have observed this problem and appraised it in different ways; for example, Jennings, Averill, Opton, and Lazarus (1971), investigating several of the more popular variables affecting heart rate–performance relationships, came to the conclusion that attention, whether altered by task demands, motor demands, uncertainty, or noxiousness of the stimuli, seems to be the common intervening variable which determines the degree of heart rate deceleration in an experimental situation. If all of these factors represent aspects of attention, it might be one of those terms which has lost its usefulness through too broad a definition. At its logical extreme, virtually any stimulus perceived by an experimental subject can be said to have attentional factors associated with it, and therefore, any concomitant heart rate change may be presumed to provide data relevant to the Lacey hypothesis. It would seem that clarification and more circumscribed and operationalized definitions of attention might help the hypothesis retain its heuristic value.

A second important element has remained obscure throughout the 12-year history of the theory. That is, how does the entire process of heart rate deceleration, baroreceptor reflex, and cortical activation begin?[9] Perusing the literature, one gets the impression that certain tasks or stimuli "reflexly" initiate heart rate acceleration or deceleration; yet, the heart rate reaction is supposed to moderate the subject's reaction to these stimuli. The temporal relationships among events and the mediational mechanisms at this input stage of the sequence have not been specified. One would suppose that an organism's initial reaction to an unfamiliar stimulus would be considerably different than that occurring later in the sequence after the subject is (presumably) more aware of the need to reject or accept the stimulus in question. In addition, it seems necessary to specify how the "appraisal mechanism," apparently a cognitive and cortical event, produces its effects on the cardiac mechanisms involved initially, before they, in turn, produce facilitating or modulating effects on the cortex. Given that the heart, itself, does not have the powers of perception, there must be some such first stage; and given the complexity of the human subject, it seems unlikely that various stimuli—especially the very complex stimuli used in many of the studies cited herein—have innate cardiac patterns associated with them. Currently, several investigators are working on this problem (see Headrick & Graham, 1969; Wilson, 1969) and the Laceys appear to be approaching the issue in their most recent work in which the heart rate deceleration is being related to the contingent negative variation EEG component (Lacey & Lacey, 1970). However, much greater detail on these formative mechanisms is needed, and a useful addition to the theory would be the elucidation of the proposed *development* of such heart rate reactions and of the results of cortical effects on cardiac responses.

---

[8] Presented in preliminary form as L. F. Delfini and J. J. Campos, Signal detection during four EKG cycle phases." Paper read at the annual meeting of the Society for Psychophysiological Research, St. Louis, October 1971.

[9] This issue was originally brought to the author's attention by Paul Obrist.

A related difficulty in understanding the hypothesis stems from the use of the term "cortical activation." It seems that at one point cortical activation occurs because heart rate decelerates. Later, the functions are reversed and heart rate decelerates because cortical activation is "needed" for a given task. Why does this cortical activation selectively facilitate sensory and sensory-motor tasks but impair performance of intellectual-associative processes? One might suspect that cortical activation would have varied effects on sensory, motor, and emotional responses and learning depending on such things as history of the organism, stimulus intensity and state or arousal level of the organism. An attempt at refinement and explication of cortical activation and its relation to other organismic variables is another area which the hypothesis could be strengthened.

A third central problem area is the instrumental role of heart rate, the implications of causality, and the degree of importance that should be attached to this one response system. Although the Laceys appear to have avoided the term "instrumental" since a cautiously worded passage in the 1959 article, subsequent offerings strongly imply that causality is being ascribed to the heart rate and blood pressure reactions. While there is little question that blood pressure through carotid sinus mechanisms *can* influence cortical excitability (this is what the neurophysiological evidence seems to indicate), whether improved reaction time performance is primarily a result of these mechanisms being activated is another issue. As we have seen, only a small proportion of the variance can be attributed to these factors on the basis of current information, and the evidence suggests mainly that those subjects who show a slowed heart rate during the foreperiod have a slightly faster reaction time—to the best of our knowledge, no study has shown that an individual can systematically *improve* his reaction time performance by learning to slow heart rate more during the foreperiod.

Heart rate is an exceedingly complex response, affected not only by the stimuli impinging on the organism (and all the possible variations in parameters of these stimuli) but also by task requirements, respiration, state of the organism, the species investigated, and changes due to repeated presentations of a stimulus (cf. Headrick & Graham, 1969; Jennings et al., 1971; Obrist et al., 1970). The Lacey hypothesis appears to have overlooked or ignored many of these important determinants of heart rate and performance variables with the result that many findings are not consistent with the notion that heart rate deceleration instrumentally facilitates performance through cortical excitation. The interaction of several of the relevant factors is demonstrated in a recent study in which task variables, preparatory intervals, attention, and noxiousness were all varied and the authors report that "all experimental manipulations appeared to contribute to cardiac deceleration . . . [Jennings et al., 1971]."

It is clear from this examination of the hypothesis that it will take an exceptionally clever research design to operationalize and differentiate the input variables, extremely precise instrumentation to evaluate the posulated mediating mechanisms, and very detailed parametric studies to relate these factors individually to the performance variables. It seems unlikely that, in the near future, we will be able to satisfactorily unravel dimensions of attention, acceptance, and pleasantness from each other or obtain an accurate measure of total baroreceptor output.

Nevertheless, the hypothesis has made some important positive contributions to the development of psychophysiological research. Perhaps the most important of these has been the role the hypothesis has played in encouraging modification of activation theory. Just as advances in neurophysiology necessitated a change of conceptualization of the Ascending Reticular Activating System from a functionally and anatomically undifferentiated network of activating cells to a markedly specialized and dynamic activating and inhibiting system, so has the Laceys' emphasis on the inverse correlation between heart rate and arousal helped to produce a modification of the earlier activation positions of Duffy (1951) and Malmo (1959). Who knows how long the "inconsistent" decelerative heart rate findings could have been overlooked had not the forceful leadership of the Laceys called attention in a very provocative and positive way to this phenomenon? In no way do these advances spell the end of activation theory or invalidate it (any more than recent knowledge of the reticular formation dispels its activating functions); rather, it is now clear that the positive correlations among automatic nervous system (ANS) measures predicted with increasing arousal do not always occur if we work within limited ranges of the arousal continuum; here, one must expect more complicated and subtle interactions among the various arousal and performance measures investigated.

Focus on the decelerative heart rate response in the context of the Laceys' hypothesis may also have made this specific response more generally acceptable or recognizable. From 8 to 10 years ago, there was a bewildering array of conflicting results on the heart rate response to different situations, including classical conditioning (cf. McDonald, Stern, & Hahn, 1963; Shearn, 1961) perhaps partly because the decelerative anticipatory heart rate response did not "fit" with a stimulus substitution view of conditioning or with the activation that would be presumed to be present during a CS. More recently, the converging trends of investigations of the orienting response and heart rate during the foreperiod of reaction time studies, improved methodology, and more careful evaluation of the heart rate response are producing growing consensus about the nature of the conditional heart rate reaction.

A further contribution of the Lacey formulation involves the emphasis it has placed on the *transactional* nature of ANS responses. As Averill,[10] among others has pointed out, the well-known conceptualization of the ANS as a primitive, unresponsive system is steeped in the psychophysiological symbolism of antiquity, and with a few exceptions (cf. Stern, 1968), it was rare to ascribe important instrumental attributes to this presumably reflexive and untalented portion of the nervous system. The Laceys' elevation of one single ANS response system to the status of a major mediating mechanism playing an important instrumental role in the performance of an organism antedated the explosion caused by Neal Miller and his colleagues upon introduction of their impressive work on instrumental control of visceral responses. Only recently, in fact, has this work presented data on the possible significance of the heart rate response as an important mediating mechanism for emotional and performance attributes of the organism (DiCara & Weiss, 1969), work which may eventually provide an important link with, and perhaps take advantage of, the theory and data generated by attempts to evaluate the hypothesis of Lacey and Lacey.

---

[10] J. R. Averill. Emotion and visceral activity: A case study in psychophysiological symbolism. Paper presented at the Society for Psychophysiological Research, Monterey, Calif., October 1969.

REFERENCES

Birren, J. E., Cardon, P. V., & Phillips, S. L. Reaction time as a function of the cardiac cycle in young adults. *Science,* 1963, **140**, 195–196.

Bonvallet, M., & Allen, M. B., Jr. Prolonged spontaneous and evoked reticular activation following discrete bulbar lesions. *Electroencephalography and Clinical Neurophysiology,* 1963, **15**, 969–988.

Bonvallet, M., Dell, P., & Hiebel, G. Tonus sympathique et activité électrique corticale. *Electroencephalography and Clinical Neurophysiology,* 1954, **6**, 119–144.

Callaway, E., & Layne, R. S. Interaction between the visual evoked response and two spontaneous biological rhythms: The EEG alpha cycle and the cardiac arousal cycle. In R. Katzman (Ed.), Sensory evoked response in man. *Annals of the New York Academy of Sciences,* 1964, **112**. (Article I)

Callaway, E., & Yeager, C. L. Relationship between reaction time and electroencephalographic alpha phase. *Science,* 1960, **132**, 1965.

Campos, J. J., & Johnson, H. J. The effects of verbalization instructions and visual attention on heart rate and skin conductance. *Psychophysiology,* 1966, **2**, 305.

Campos, J. J., & Johnson, H. J. Affect, verbalization and directional fractionation of autonomic responses. *Psychophysiology,* 1967, **3**, 285.

Chase, W. G., Graham, F. K., & Graham, D. T. Components of the heart rate response in anticipation of reaction time and exercise tasks. *Journal of Experimental Psychology,* 1968, **76**, 642–648.

Davis, R. C. Response patterns. Transactions of the New York Academy of Science, 1957, **19**, 731–739. (Series II)

Davis, R. C., & Buchwald, A. M. An exploration of somatic response patterns: Stimulus and sex differences. *Journal of Comparative Physiological Psychology,* 1957, **50**, 44–52.

Davis, R. C., Buchwald, A. M., & Frankmann, R. W. Autonomic and muscular responses, and their relation to simple stimuli. *Psychological Monographs,* 1955, **69**(20, Whole No. 405).

Dell, P. C. Humoral effects on the brain stem reticular formations. In H. H. Jasper, L. D. Proctor, R. S. Knighton, W. C. Noshay, & R. T. Costello (Eds.), *Reticular formation of the brain.* Boston: Little, Brown & Co., 1957.

Dell, P., Bonvallet, M., & Hugelin, A. Tonus sympathique, adrenaline et controle reticulaire de la motricite spinale. *Electroencephalography and Clinical Neurophysiology,* 1954, **6**, 599–618.

DiCara, L. V., & Weiss, J. M. Effect of heart-rate learning under curare on subsequent noncurarized avoidance learning. *Journal of Comparative and Physiological Psychology,* 1969, **69**, 368–374.

Duffy, E. The concept of energy mobilization. *Psychological Review,* 1951, **58**, 30–40.

Edwards, D. C., & Alsip, J. E. Intake-rejection, verbalization, and affect: Effects on heart rate and skin conductance. *Psychophysiology,* 1969, **6**, 6–12.

Elliot, R. The significance of heart rate for behavior: A critique of Lacey's hypothesis. *Journal of Personality and Social Psychology,* 1972, **22**, 398–409.

Graham, F. K., & Jackson, J. C. Arousal systems and infant heart rate responses. In L. P. Lipsitt & H. W. Reese (Eds.), *Advances in child development and behavior.* New York: Academic Press, in press.

Headrick, M. W., & Graham, F. K. Multiple-component heart rate responses conditioned under paced respiration. *Journal of Experimental Psychology,* 1969, **79**, 486–494.

Hein, P. L. Heart rate conditioning in the cat and its relationship to other physiological responses. *Psychophysiology,* 1969, **5**, 455–464.

Jennings, J. R., Averill, J. R., Opton, E. M., & Lazarus, R. S. Some parameters of heart rate change: Perceptual versus motor task requirements, noxiousness and uncertainty. *Psychophysiology,* 1971, **7**, 194–212.

Johnson, H. J., & Campos, J. J. The effect of cognitive tasks and verbalization instructions on heart rate and skin conductance. *Psychophysiology,* 1967, **4**, 143–150.

Lacey, J. I. Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. *Research in psychotherapy.* Vol. 1. Washington, D.C., American Psychological Association, 1959.

Lacey, J. I. Somatic response patterning and stress: Some revisions of activation theory. *Psychological stress: Issues in research.* New York: Appleton-Century-Crofts, 1967.

Lacey, J. I., Kagan, J., Lacey, B. C., & Moss, H. A. The visceral level: Situational determinant and behavioral correlates of autonomic response patterns. In P. H. Knapp (Ed.), *Expression of the emotions in man.* New York: International Universities Press, 1963.

Lacey, J. I., & Lacey, B. C. The relationship of resting autonomic activity to motor impulsivity. *Research Publications: Association for Research in Nervous & Mental Disease,* 1958, **36**, 144–209.

Lacey, J. I., & Lacey, B. C. Some autonomic-central nervous system interrelationships. In P. Black (Ed.), *Physiological correlates of emotion.* New York: Academic Press, 1970.

Lewis, M., Kagan, J., Campbell, H., & Kalafat, J. The cardiac response as a correlate of attention in infants. *Child Development,* 1966, **37**, 63–71.

Malmo, R. B. Activation: A neuropsychological dimension. *Psychological Review,* 1959, **66**, 367–386.

McDonald, D. G., Stern, J. A., & Hahn, W. W. Classical heart rate conditioning in the rat. *Journal of Psychosomatic Research,* 1963, **7**, 97–106.

Nakao, H., Ballim, H. M., & Gellhorn, E. The role of the sinoaortic receptors in the action of adrenaline, noradrenaline and acetylchoine on the cerebral cortex. *Electroencephalography and Clinical Neurophysiology,* 1956, **8**, 413–420.

OBRIST, P. A. Cardiovascular differentiation of sensory stimuli, *Psychosomatic Medicine,* 1963, 25, 450–459.

OBRIST, P. A., WEBB, R.A., & SUTTERER, J. R. Heart rate and somatic changes during aversive conditioning and a simple reaction time task. *Psychophysiology,* 1969, 5, 696–723.

OBRIST, P. A., WEBB, R. A., SUTTERER, J. R., & HOWARD, J. L. Cardiac deceleration and reaction time: An evaluation of two hypotheses. *Psychophysiology,* 1970, 6, 695–706.

SAXON, S. A. Detection of near threshold signals during four phases of cardiac cycle. *Alabama Journal of Medical Sciences,* 1970, 7.

SHEARN, D. Does the heart learn? *Psychological Bulletin,* 1961, 68, 452–458.

STEELE, W. G., & KOONS, P. B., JR. Cardiac response to mental arithmetic under quiet and white noise distraction. *Psychonomic Science,* 1968, 11, 273–274.

STEELE, W. G., & LEWIS, M. A longitudinal study of the cardiac response during problem solving task and its relationship to general cognitive function. *Psychonomic Science,* 1968, 11, 275.

STERN, J. A. Toward a developmental psychophysiology: My look into the crystal ball. *Psychophysiology,* 1968, 4, 403–420.

THOMPSON, L. W., & BOTWINICK, J. Stimulation in different phases of the cardiac cycle and reaction time. *Psychophysiology,* 1970, 7, 57–65.

WEBB, R. A., & OBRIST, P. A. The physiological concomitants of reaction time performance as a function of preparatory interval and preparatory interval series. *Psychophysiology,* 1970, 6, 389–403.

WILSON, R. S. Cardiac response: Determinants of conditioning. *Journal of Comparative and Physiological Psychology,* 1969, 68, 1–23.

(Received August 2, 1971)

Manuscripts Accepted for Publication in the *Psychological Bulletin*

Error Rates for Multiple Comparison Methods: Some Evidence Concerning the Misleading Conclusions of Petrinovich and Hardyck. H. J. Keselman (Department of Psychology, University of Manitoba, Winnipeg, Manitoba, Canada) and Larry E. Toothaker.

Rejoinder to Morgan. Wallace Wilkins (Department of Psychology, University of Miami, P.O. 8185, Coral Gables, Florida 33124).

Computational Procedures for Estimating Magnitude of Effect for Some ANOVA Designs. David H. Dodd (Department of Psychology, University of Utah, Salt Lake City, Utah 84112) and Roger F. Schultz, Jr.

ERRATUM

In the article, "Septum and Behavior: A Review," by P. A. Fried in the October 1972 issue, line 19 of the second column of page 292 should read as follows:

. . . separate paper (Fried, 1972). One . . . .

Psychological Bulletin
1973, Vol. 79, No. 1, 71-72

TESTING IF TWO MEASURING PROCEDURES MEASURE THE SAME DIMENSION[1]

FREDERIC M. LORD[2]

Educational Testing Service

A convenient statistical technique is described for testing the hypothesis that two sets of measurements differ only because of errors of measurement and because of differing origins and units of measurement.

This note is concerned with testing the hypothesis that two sets of measurements differ only because of (*a*) errors of measurement, (*b*) differing units of measurement, and (*c*) differing arbitrary origins for measurement. It describes a convenient statistical technique that, unnoticed, has become available (Villegas, 1964). It will probably be preferred to other techniques used for this purpose (Forsyth & Feldt, 1969, 1970; Lord, 1957; McNemar, 1958), though each has advantages and disadvantages.

STATISTICAL PROCEDURE

The necessary raw data consist of $r \geq 2$ replicate measurements (indexed by the subscript k) on each of N people (indexed by a) by each of two measuring instruments or procedures (indexed by g or h). For each instrument or procedure separately, compute the usual among-persons and within-persons sums of squares. Also, compute the corresponding sums of cross products between instruments (or procedures). Let W denote the 2×2 matrix of within-persons sums of squares and cross products, and let A denote the corresponding 2×2 among-persons matrix. In a standard notation, the element of W in column g and row h is

$$\sum_{a=1}^{N} \sum_{k=1}^{r} (x_{gak} - x_{ga.})(x_{hak} - x_{ha.}).$$

A typical element of A, similarly, is

$$r \sum_{a=1}^{N} (x_{ga.} - x_{g..})(x_{ha.} - x_{h..}).$$

Let F_p denote the $1 - p$ percentile of the F distribution with N and $N(r - 1)$ degrees of freedom, for the numerator and denominator, respectively. Compute the matrix:

$$M = (r - 1)A - F_pW.$$

From the product of the two diagonal terms of M subtract the product of the two off-diagonal terms, thus obtaining the value of the determinant $|M|$. If the determinant is positive and if both diagonal terms are also positive, then the F test rejects at significance level $< p$ the hypothesis H_0 stated at the beginning of this paper. Otherwise, the hypothesis is not rejected. This same conclusion may be stated more compactly as follows: H_0 is rejected if and only if the matrix M is positive definite.

DISCUSSION

The assumptions required for the validity of the foregoing statistical procedure and conclusions are as follows. Within each replication, the errors of measurement for instruments g and h for each person always have a bivariate normal distribution. The mean error is always taken to be zero. No restrictive assumptions are made about the variance or correlation parameters of this distribution, except that the variances are positive. In particular, the errors for g and h may be correlated within a replication and within a person. The bivariate distribution is the same for each replication and for each individual measured. For two replications or two persons,

---

[1] This research was sponsored in part by the Personnel and Training Research Programs, Psychological Sciences Division of the Office of Naval Research under Contract No. N00014-69-C-0017 and Contract Authority Identification NR 150-303 and by Educational Testing Service. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Requests for reprints should be sent to Frederic M. Lord, Division of Psychological Studies, Educational Testing Service, Princeton, New Jersey 08540.

TABLE 1

Data for Speeded and Unspeeded Vocabulary Tests

Group	Unspeeded test SS	Sum of products	Speeded test SS
Among persons	93524.56	76176.30	111220.58
Within persons	18533.00	−547.00	15403.50
Total	112057.56	75629.30	126624.08

the errors are always uncorrelated with each other.

The unknown relationship between the two origins mentioned in H_0, also that between the two units of measurement, represent "nuisance parameters" for the purpose of making a significance test of H_0. Villegas (1964) shows that when (and only when) M is positive definite, there is no set of values for these (unknown) nuisance parameters such that an appropriate variance ratio calculated from them and from the data will lie below the p level of significance.

The significance level in the foregoing statement is precisely p. However, when the entire procedure is considered as a test for H_0, it is known only (assuming that the replicate measurements are not perfectly correlated) that the significance level of the test will be less than p. Thus, the significance test is conservative in the sense that H_0 will be rejected somewhat less often than the p value would indicate.[3] It is important in practice that the replications meet the assumptions stated. Any practice effect between replications that increases the within-persons sums of squares, for example, will tend to decrease the chance of rejecting H_0.

Numerical Example

In the numerical example, measuring instruments g are unspeeded 15-item vocabulary tests; measuring instruments h are highly speeded 75-item vocabulary tests. The $r = 2$ parallel forms of each test were administered to each of $N = 649$ examinees. The raw data are the same as those used in a previous numerical example (Lord, 1957, pp. 210–212). The required sums of squares and products are shown in Table 1.

The .05 significance level of F for 649 and 649 degrees of freedom is 1.13. The matrix M is easily found (to the nearest integer) to be

$$M = \begin{bmatrix} 72582 & 76794 \\ 76794 & 93815 \end{bmatrix}.$$

The determinant is positive, so H_0 is rejected. This agrees with the conclusion reached previously by large-sample methods (Lord, 1957) under somewhat different assumptions.

REFERENCES

Forsyth, R. A., & Feldt, L. S. An investigation of empirical sampling distributions of correlation coefficients corrected for attenuation. *Educational and Psychological Measurement*, 1969, **29**, 61–72.

Forsyth, R. A., & Feldt, L. S. Some theoretical and empirical results related to McNemar's test that the population correlation coefficient corrected for attenuation equals 1.0. *American Educational Research Journal*, 1970, **7**, 197–207.

Lord, F. M. A significance test for the hypothesis that two variables measure the same trait except for errors of measurement. *Psychometrika*, 1957, **22**, 207–220.

McNemar, Q. Attenuation and interaction. *Psychometrika*, 1958, **23**, 259–266.

Saw, J. G. A conservative test for the concurrence of several regression lines and related problems. *Biometrika*, 1966, **53**, 272–275.

Villegas, C. Confidence region for a linear relation. *The Annals of Mathematical Statistics*, 1964, **35**, 780–788.

(Received August 23, 1971)

---

[3] The author is indebted to Murray Aitkin and to Leon Gleser for pointing out and clarifying this distinction. See also Saw (1966, Sections 2–3).

VOL. 79, NO. 2 FEBRUARY 1973

Psychological Bulletin

Copyright © 1973 by the American Psychological Association, Inc.

SELF-DISCLOSURE:

A LITERATURE REVIEW[1]

PAUL C. COZBY[2]

University of Minnesota

Literature relating to the verbal disclosure of information about oneself is reviewed. First, the role of self-disclosure as a personality attribute is considered. It is suggested that the failure of researchers to find consistent relationships between self-disclosure and other personality variables is due to the poor development of paper-and-pencil measures of disclosure which themselves do not relate to actual disclosure. It is suggested that researchers interested in personality correlates of self-disclosure employ behavioral measures of disclosure. The role of self-disclosure in interpersonal relationships is considered, and the principle variables that appear to facilitate disclosure are reviewed. However, it is suggested that factors inhibit intimate disclosure, such as needs for privacy, have not been adequately treated in disclosure research. The role of self-disclosure in experimenter–subject, interviewer–interviewee, and therapist–client relationships is reviewed. Finally, the review considers the relationship between self-disclosure and nonverbal aspects of communication.

Self-disclosure may be defined as any information about himself which Person A communicates verbally to a Person B. The term "self-disclosure" has been used by Jourard (e.g., 1964), although other terms such as "verbal accessibility" (e.g., Polansky, 1965), and "social accessibility" (e.g., Rickers-Ovsiankina, 1956) have been used to describe the same concept. "Self-disclosure" is used in the present paper because (*a*) it is the most commonly used term in the literature and (*b*) the term refers to both a personality construct and a process which occurs during interaction with others.

MEASURING SELF-DISCLOSURE

Jourard's Self-Disclosure Questionnaire

The most widely used instrument to assess individual differences in self-disclosure has been Jourard's Self-Disclosure Questionnaire (JSDQ). The initial instrument, described by Jourard and Lasakow (1958), consisted of 60 items—10 items in each of six content areas: attitudes and opinions, tastes and interests, work (or studies), money, personality, and body. Subjects responded to each item by indicating the extent to which the information has been revealed to four target persons: Mother, father, best opposite-sex friend, and best same-sex friend. Items were scored as 0—no disclosure to the target person, 1—disclosure only in general terms, or 2—full and complete disclosure about the item.

Pederson and Higbee (1968) point out that subsequent research has employed variations of the JSDQ that differ on a number of dimensions, including length of questionnaire, target persons, instructions, and nature of the

---

[1] Preparation of this review was supported by a grant from the University of Minnesota Agriculture Experiment Station to Paul C. Rosenblatt. The author would like to thank Paul C. Rosenblatt, Ellen Berscheid, and Karl E. Weick for their helpful comments on an earlier version of the manuscript, and Sidney M. Jourard and Dalmas A. Taylor for making unpublished materials available.

[2] Requests for reprints should be sent to Paul C. Cozby who is now at the Department of Psychology, California State University, Fullerton, Fullerton, California 92634.

705 (5)

items. Despite such differences (which are described at length by Pederson & Higbee, 1968), there has been a tacit assumption in the literature that the various measures are equivalent.

Validity of the JSDQ. Pederson and Higbee (1968) obtained evidence for convergent and discriminant validity of the 60- and 25-item JSDQ by means of multitrait–multimethod matrices (Campbell & Fiske, 1959), but there was also evidence for variation between the two methods of measuring disclosure. The JSDQ appears to be independent of intelligence (Jourard, 1961d; Halverson & Shore, 1969; Taylor, 1968) providing evidence for discriminant validity of the JSDQ.

Jourard (1961c) reported that productivity on the group Rorschach correlated .37 ($p < .05$) with total score on the JSDQ, and interpreted the finding as supportive of the validity of the instrument.

However, there is little evidence for the predictive validity of the JSDQ. Researchers have been unable to find a relationship between the JSDQ and actual disclosure in a situation (Ehrlich & Graeven, 1971; Himelstein & Kimbrough, 1963; Lubin & Harrison, 1964; Vondracek, 1969a, 1969b) or ratings of actual disclosure made by peers (Himelstein & Lubin, 1965; Hurley & Hurley, 1969). Pederson and Breglio (1968b) did find that amount and intimacy of disclosure on written self-descriptions were correlated with total scores on the 60-item JSDQ, but not on the 25-item JSDQ. Burhenne and Mirels (1970) found that rated disclosure on written self-descriptions correlated exactly .00 with the JSDQ.

It is clear that the JSDQ does not accurately predict actual self-disclosure. The explanation perhaps lies in the fact that scores on the JSDQ reflect subjects' past history of disclosure to parents and persons who are labeled "best same-sex friend" and "best opposite-sex friend." When actual disclosure is measured, the subject is disclosing to an experimenter or to peers whom the subject has never met.

Jourard and his students have recently begun using a 40-item questionnaire which asks subjects to indicate what they have disclosed to someone in the past, and what they would be willing to disclose to a stranger of the same sex (see Jourard, 1969, Appendix 1, for a description of the questionnaire). This questionnaire has been shown to predict actual disclosure in studies by Drag[3] and Jourard and Resnick (1970). These studies also indicate that situational variables may outweigh individual differences in disposition to disclose.

The problem of the validity of the JSDQ is discussed further after the review of attempts to relate personality variables to self-disclosure.

Other Measures of Self-Disclosure

Several investigators have developed other measures of self-disclosure. Since these measures have not been widely used, or were developed for specific types of subjects, they are discussed only briefly.

1. West and Zingle (1969) describe a Self-Disclosure Inventory for Adolescents.
2. Vondracek and Vondracek (1971) have developed a system for scoring self-disclosure by preadolescents in interview situations.
3. The Social Accessibility Scale consisting of 25 items has been described by Rickers-Ovsiankina (1956), and a 50-item version is described by Rickers-Ovsiankina and Kusmin (1958). The scale differs from the JSDQ in that (*a*) subjects are instructed to indicate what they *would* disclose rather than what they *have* disclosed and (*b*) the target persons are a stranger, an acquaintance, and a best friend. Pederson and Higbee (1968) correlated the 60- and 25-item JSDQ with the 50-item Social Accessibility Scale, and concluded that the two measures should not be considered equivalent, as has been done previously (cf. Jourard, 1964).
4. Polansky's (1965) concept of "verbal accessibility" has been measured by the incomplete sentence method, degree of agreement with such statements as "the really smart guys keep their opinions to themselves," and reports of caseworkers (Polansky & Brown, 1967).

---

[3] L. R. Drag. Experimenter-subject interaction: A situational determinant of differential levels of self-disclosure. Unpublished master's thesis, University of Florida, 1968.

5. Taylor and Altman (1966) scaled 671 statements for intimacy value and topical category. Although these statements do not comprise a disclosure questionnaire, the statements can be used to construct a questionnaire. The previously discussed measure currently being used by Jourard (1969) uses the Taylor and Altman statements. A 144-item questionnaire using the statements, which uses "best friend" as the only target person, correlates .25 ($p < .05$) with actual disclosure in an interview situation (Vondracek & Marshall, 1971). Although significant, this correlation is not impressive. It would be interesting to know if such a correlation would be raised by including "stranger" as a target person. The statements may also be used as stimulus materials for experimental manipulations, and as a means of scoring speech content for self-disclosure.

PARAMETERS OF SELF-DISCLOSURE

The basic parameters of self-disclosure are (*a*) breadth or amount of information disclosed, (*b*) depth or intimacy of information disclosed, and (*c*) duration or time spent describing each item of information.

In actual disclosure, intimacy and duration of disclosure appear to be partially independent, with a correlation of .42 between the two variables (Vondracek, 1969a). There is an inverse relationship between amount and intimacy of disclosure such that individuals disclose less about more intimate topics (Altman & Haythorn, 1965; Fitzgerald, 1963; Hood & Back, 1971; Jourard & Lasakow, 1958; Taylor, 1968; Taylor, Altman, & Sorrentino, 1969; Tuckman, 1966), and the effect appears to hold across cultures (Brown, 1965; Melikian, 1962).

Little work has been done on the content of information disclosed, although there is some research on the positive or negative aspects of the information. Levin and Gergen (1969) found that individuals disclose more positive aspects of themselves when trying to be ingratiating than when trying to be honest. Wives tend to communicate more unpleasant feelings, including anxieties and worries, but not more pleasant feelings than do husbands (Katz, Goldston, Cohen, & Stucker, 1963; Levinger & Senn, 1967).

SELF-DISCLOSURE AS A PERSONALITY CONSTRUCT

Self-Disclosure and Family Patterns

One question one may ask when dealing with personality variables is what sorts of child-rearing practices and family interaction patterns are associated with variations in the personality variable in adulthood.

Dimond and Munz (1967) found and later replicated (Diamond & Hellkamp, 1969), that later borns show higher self-disclosure scores than firstborns. They concluded that while firstborns may have greater affiliative needs than later borns, they are not as adept at satisfying them, at least by establishing close relationships with others.

Jourard (1961b) found that Jewish males were significantly higher in disclosure than Baptists, Methodists, and Catholics, none of which differed significantly from one another. However, Low[4] did not find a difference between Jewish and Protestant groups. Cooke[5] obtained a correlation of .22 ($p < .05$) between disclosure to parents and religious behavior (e.g., frequency of church attendance) of protestants. Jennings (1971) did not find a relationship between disclosure and whether religious beliefs are "liberal" or "conservative." Presumably, any effects of religious background on disclosure are mediated by a closer family orientation.

A recent study by Pederson and Higbee (1969b) found that disclosure to parents was correlated with subjects' ratings of parents on such adjectives as close, warm, friendly, and accepting. In addition, it was found that females who rated the mother as cold, distrustful, and selfish tended to score high on the Social Accessibility Scale which measures willingness to disclose to strangers, acquaintances, and/or best friends. Thus, it seems possible that family relationships are more important in determining who a person discloses to rather than whether or not the person will be a high discloser. Such an interpretation is supported by Doster and Strickland's (1969) finding that, while, in general, high disclosers perceive their parents as more nurturant than low disclosures, subjects from the low-nurturant homes disclose more to friends than parents while the reverse is true with subjects from high-nurturant families.

---

[4] A. B. Low. Sex and religious differences in self-disclosure as determined by a revised method of scoring the "Jourard Self-Disclosure Questionnaire." Unpublished master's thesis, University of Florida, 1964.

[5] T. F. Cooke. Interpersonal correlates of religious behavior. Unpublished doctoral dissertation, University of Florida, 1962.

Sex, Race, and Cultural Factors in Self-Disclosure

Sex differences. Jourard and Lasakow (1958) reported that females have higher disclosure scores than males. This effect has been replicated in numerous investigations (Dimond & Munz, 1967; Himelstein & Lubin, 1965; Hood & Back, 1971; Jourard & Landsman, 1960; Jourard & Richman, 1963; Pederson & Breglio, 1968a; Pederson & Higbee, 1969a). Jourard (1964, Ch. 6) attributed considerable importance to the obtained sex differences. The low disclosure of males was seen to be directly associated with less empathy, insight, and a shorter life-span than females. However, a number of studies have reported no sex differences in self-disclosure (Dimond & Hellkamp, 1969; Doster & Strickland, 1969; Low, (see Footnote 4); Plog, 1965; Rickers-Ovsiankina & Kusmin, 1958; Vondracek & Marshall, 1971; Weigel, Weigel, & Chadwick, 1969; Zief[6]).

Jourard (1964) and Plog (1965) have suggested that conflicting findings concerning sex differences may be the result of samples from different geographical areas with a concomitant difference in sex role expectations. Sex differences would be found in the southern United States, but not in the northern and western United States. However, an examination of the studies which tested for sex differences yields no consistent pattern which would allow the conflicting results to be explained by differences in geographic locale or type of instrument used.

The fact that no study has reported greater male disclosure may be indicative of actual sex differences. The nature of any sex differences might be found if researchers were to pay greater attention to the types of items which reliably discriminate between males and females, and types of situations in which males and females would or would not differ in disclosure output. For instance, Pederson and Breglio (1968a), using written self-descriptions, found that females did not use more words to describe themselves than males, but they disclosed more intimate information about themselves than males.

Race, ethnic group, and social class factors. Jourard and Lasakow (1958) reported less disclosure by blacks than by whites. This finding has been replicated by Dimond and Hellkamp (1969), and by Littlefield[7] who also reported less disclosure by Mexican-Americans than by blacks. There is some evidence that racial differences in self-disclosure may be due to social class factors. Jaffee and Polansky (1962) found no differences in disclosure between lower-class blacks and lower-class whites. Mayer (1967) had reported that middle-class women disclose more about their marital difficulties than do working-class women.

Tulkin (1970) presents data from an ongoing study which indicate that middle-class mothers initiate a greater number of verbalizations and respond to a greater percentage of the child's vocalizations than lower-class mothers. These results suggest that the reinforcement patterns that lead to variations in self-disclosure in adulthood may be operating as soon as the infant begins to babble.

Cross-cultural findings. Lewin (1948) observed that Americans disclose a great deal about themselves and make friends easily, but do not develop highly intimate relationships, while Germans don't disclose much about themselves to others in general, but they do become quite intimate with a few others. Plog (1965) obtained evidence for the hypothesis that Americans disclose more than Germans. However, the hypothesis that Germans would disclose more than Americans to a close friend of the same or opposite sex was not supported.

---

[6] R. M. Zief. Values and self-disclosure. Unpublished honors' thesis, Harvard University, 1962.

[7] R. P. Littlefield. An analysis of the self-disclosure patterns of ninth grade public school children in three selected subcultural groups. Unpublished doctoral dissertation, Florida State University, 1968.

TABLE 1

STUDIES WHICH EXAMINED THE RELATIONSHIP BETWEEN SELF-DISCLOSURE AND MENTAL HEALTH

Author and date	Subjects			Measure of mental health
	Sex	Age	*n*	
Positive relationships between disclosure and mental health				
Mayo (1968)	F	Adult	80	Neurotic inpatients compared with normals
Taylor, Altman, & Frankfurt (see Footnote 8)	M	17–20[a]	100	MMPI *Hy*, *Hs*, *Sc*, *Pd*
Pederson & Higbee (1969a)	F	CS	51	Pederson Personality Inventory Cycloid disposition scale
Negative relationships between disclosure and mental health				
Pederson & Higbee (1969a)	M	CS	56	Pederson Personality Inventory Cycloid disposition scale
Pederson & Breglio (1968a)	F	CS	26	Cycloid disposition scale
Truax & Wittmer (in press)	M/F	CS	89	MMPI adjustment profile
Persons & Marks (1970)	M	18–22[b]	81	MMPI *F*, *Pd*, *Pa*, *Sc*
Cozby (1972)	F	CS	30	Ratings of hypothetical stranger disclosing at low, medium, and high levels of intimacy
No relationship between disclosure and mental health				
Stanley & Bownes (1966)	M/F	CS	137	Maudsley Personality Inventory neuroticism scale
Pederson & Breglio (1968a)	M	CS	26	Pederson Personality Inventory Cycloid disposition scale
Jourard (1964, p. 181)	—	CS	—	Applicants for clinical services

Note. Abbreviations: CS = college students.
[a] Naval recruits.
[b] Prison inmates.

Self-Disclosure and Mental Health

Jourard (1959a) argued that the ability to allow one's real self to be known to at least one "significant" other is a prerequisite for a healthy personality. Concerned with the concept of self-actualization (Maslow, 1954), Jourard proposed that low disclosure is indicative of a repression of self and an inability to grow as a person. Jourard (1964) elaborated on the self-disclosure–mental-health relationship, and suggested that the relationship between the two variables is curvilinear. Although this brief summary does little justice to Jourard's (1959a, 1963, 1964) views, it serves as an introduction to research investigating the relationship between self-disclosure and mental health.

Jourard's writings indicate that disclosure should be negatively related to "clinical" maladjustment and also positively to "positive" mental health (e.g., self-actualization). Table 1 summarizes the studies which have looked at the former type of mental health. With regard to the studies that support Jourard's initial hypothesis, it should be pointed out that the magnitude of the correlations are small: .28 in the Pederson and Higbee (1969a) study, and .18 to .34 in the Taylor, Altman, and Frankfurt[8] study, and these authors report that an attempted replication yielded even lower correlations.

Although it is difficult to identify the criteria that should be used to judge whether a

---

[8] D. A. Taylor, I. Altman, & L. P. Frankfurt. Personality correlates of self-disclosure. Unpublished manuscript, 1965.

TABLE 2

STUDIES WHICH EXAMINED SELF-DISCLOSURE AND "POSITIVE" MENTAL HEALTH

Author & date	Subjects			Measure of mental health	Relationship found
	Sex	Age	*n*		
Vargas (see Footnote 9)	M	CS	277	Positive experiencing and behaving scale	positive
Halverson & Shore (1969)	M/F	Adult[a]	53	Conceptual complexity, interpersonal flexibility, general adaptibility	positive
Fitzgerald (1963)	F	CS	300	Self-esteem index	*ns*
Doster & Strickland (1969)	M/F	CS	110	Marlowe-Crowne Social Desirability scale	*ns*
Burhenne & Mirels (1970)	F	CS	56	Marlowe-Crowne Social Desirability scale	*ns*

Note. Abbreviations: CS = college students.
[a] Peace Corps trainees.

study is testing the hypothesis that disclosure is related to positive mental health (cf. Smith, 1959), the studies listed in Table 2 seem at least somewhat appropriate. The Vargas[9] study which measured self-actualization is perhaps the most appropriate. However, our confidence in such a relationship would be increased by a study relating disclosure to a more standardized measure of self-actualization such as the Personal Orientation Inventory.

Is self-disclosure related to mental health? Virtually every type of relationship has been reported in the literature, yet no correlation reported has been greater than .50, and most are much lower. The curvilinear relationship suggested by Jourard (1964) seems intuitively compelling for several reasons. The individual who never discloses may be unable to establish close relationships with others. A large portion of his self may be seen as threatening and is repressed (there is evidence that repressors talk less than sensitizers—Kaplan, 1967; Axtell & Cole, 1971—but it is not known whether repressors are also less intimate). In contrast, the individual who characteristically discloses a great deal about himself, not just to someone close such as a parent or spouse, but to anyone, may be perceived by others as maladjusted and also unable to relate to others because of a preoccupation with his own self. The medium discloser may disclose a great deal to someone who is very close, and generally maintains a moderately close relationship with others (i.e., not too close to be offensive, but close enough to establish meaningful social bonds). If the relationship between self-disclosure and mental health is actually curvilinear, there are two possible reasons for the conflicting results. First, none of the studies previously cited reported computation of a correlation ratio. Significant departures from linearity thus may have gone undetected. It is also possible that true linear relationships were obtained because of biases in the sample. A positive relationship between disclosure and mental health would be found with samples that contained no high disclosers, while a negative relationship would be found in a sample that contained no low disclosers. The author would suggest the following hypotheses: Persons with positive mental health (given that they can be identified) are characterized by high disclosure to a few significant others and medium disclosure to others in the social environment. Individuals who are poorly adjusted (again assuming a suitable identification can be made) are characterized by either high or low disclosure to virtually everyone in the social environment. Future research should help to clarify this rather confused aspect of self-disclosure.

Personality Correlates of Self-Disclosure

Studies which have examined personality correlates of disclosure are shown in Table 3. Like the previously cited studies on disclosure and mental health, correlations involving self-disclosure scores and various per-

---

[9] R. Vargas. A study of certain personality characteristics of male college students who report frequent positive experiencing and behaving. Paper presented at the meeting of the Southeastern Psychological Association, New Orleans, February 1969.

TABLE 3

STUDIES WHICH EXAMINED PERSONALITY CORRELATES OF SELF-DISCLOSURE

Author & date	Subjects			Personality measure used	Relationship found
	Sex	Age	n		
			Femininity		
Taylor, Altman, & Frankfurt (see Footnote 8)	M	17–20^{a}	100	MMPI *Mf*	positive
				Guilford-Zimmerman	positive
Swensen (see Footnote 11)	—	—	—	Guilford-Zimmerman	positive
Pederson & Breglio (1968a)	M/F	CS	52	Gough femininity	*ns*
Pederson & Higbee (1969a)	M/F	CS	107	Gough femininity	*ns*
			Authoritarianism		
Taylor, Altman, & Frankfurt (see Footnote 8)	M	17–20^{a}	100	Rokeach Dogmatism	positive
Halverson & Shore (1969)	M/F	Adult^{b}	53	F scale	negative
Worthy, Gary, & Kahn (1969)	F	CS	48	Rokeach Dogmatism	*ns*
			Sociability and Extraversion		
Taylor, Altman, & Frankfurt (see Footnote 8)	M	17–20^{a}	100	Edwards Personal Preference Schedule affiliation, succorance, nurturance, heterosexuality	positive
				MMPI *Si*	negative
Tuckman (1966)	M	17–24^{a}	299	Systems of integrative complexity	positive
Taylor & Oberlander (1969) Experiment 1	M	$\bar{X} = 18.5^{a}$	42	Accuracy in a facial recognition task	positive
Swensen (see Footnote 11)	—	—	—	Maudsley Personality Inventory extraversion	positive
Swensen (see Footnote 11)	—	—	—	Guilford-Zimmerman social interest—sociability, ascendance—social boldness	positive
Frankfurt (see Footnote 10)	F	CS	60	Guilford-Zimmerman sociability, thoughtfulness, personal relations	*ns*
Pederson & Breglio (1968a)	M/F	CS	52	Pederson Personality Inventory extraversion scale	*ns*
Pederson & Higbee (1969a)	M/F	CS	107	Pederson Personality Inventory extraversion scale	*ns*
			College achievement		
Powell & Jourard (1963)	M/F	CS	80	Grades	*ns*
Jourard (1961d)	F	CS	23	Grades in nursing courses	positive
				Grades in nonnursing courses	*ns*
			Interpersonal trust		
Vondracek & Marshall (1971)	M/F	CS	106	Rotter Interpersonal Trust scale	*ns*

Note. Abbreviations: CS = college students.

[a] Naval recruits.

[b] Peace Corps trainees.

sonality measures have been generally low and often contradictory.

The pattern of results involving measures of sociability and extraversion do seem consistent, and perhaps one may have confidence in a positive relationship between disclosure and extraversion. However, the studies by Frankfurt[10] and Pederson and his associates are somewhat contradictory.

It is clear that the personality characteristics associated with self-disclosure are not well understood. Altman and Taylor (1973) feel that it is unrealistic to expect to find specific trait-disclosure relationships. Their approach is to explore the relationship between personality and self-disclosure in the context of specific relationships and settings, and is considered below.

It is also likely that the use of the JSDQ is a factor in the generally low correlations between disclosure and various personality traits (including mental health). It was previously mentioned that the JSDQ does not appear to predict actual disclosure to others. The JSDQ may be best interpreted as a measure of past history of disclosure. Some evidence of the subject's present disposition to disclose may be inferred from the subject's reports of disclosure to friends. Swensen[11] has reported that the highest personality trait-JSDQ correlations involve peers as target persons. Perhaps a more sensitive measure of disposition to disclose would be subjects' willingness to disclose to an acquaintance, a stranger, or the experimenter.

The relationship between disclosure and the Marlowe-Crowne Social Desirability Scale (SDS) is an example of the problems associated with the use of the JSDQ. Doster and Strickland (1969) found no relationship between the JSDQ and SDS. However, Burhenne and Mirels (1970) noted that the social-desirability–approval motive should be related to self-disclosure since high-SDS subjects protect their vulnerable self-esteem by being conforming, submissive, conventional, and generally self-protective. Burhenne and Mirels rated the disclosure of subjects' written self-disclosure and found a correlation of $-.41$ ($p < .01$) between disclosure and the SDS, while the JSDQ correlated only $-.03$ with the SDS.

It must be concluded that continued use of the JSDQ will only perpetuate the confusion that already exists in the literature. Alternate methods of measuring self-disclosure include written self-descriptions, and Jourard's (1969) previously discussed questionnaire since there is evidence for the predictive validity of this measure.

The author, however, feels that it would be a mistake to continue the collection of correlations between personality trait measures and self-disclosure questionnaires. Instead, self-disclosure should be measured behaviorally and used as the dependent variable. An example of this type of procedure is provided by a study by Axtell and Cole (1971), who classified subjects as repressors, sensitizers, or neutrals, and measured the amount of time subjects spent discussing either positive or negative aspects of themselves. It should be expected that any study using such a procedure would yield results which would be considerably less equivocal than studies employing any of the poorly developed self-disclosure questionnaires. The procedure would also allow the introduction of situational variables into the design, and the development of better questionnaires by administering the measure to subjects and correlating scores with subjects' actual disclosure.

SELF-DISCLOSURE IN INTERPERSONAL RELATIONSHIPS

Social Exchange and Penetration Processes

In this section, the role of self-disclosure in interpersonal relationships is considered. Self-disclosure is placed within the context of social exchange and social penetration theory. According to social penetration theory (Altman & Taylor, 1973; Taylor, Altman, & Sorrentino, 1969), "the growth of an interpersonal relationship is hypothesized to be a joint result of interpersonal reward/cost factors, personality characteristics, and situational determinants [Taylor et al., 1969, p. 325]." Relationships proceed from nonintimate to intimate areas of exchange. The rate and amount of movement from nonintimate to intimate areas of exchange (which includes verbal disclosure, type of activities engaged in, and nonverbal communication) is determined by reward/cost factors of past, present, and projected future exchanges. Social penetration theory emphasizes reward/cost factors, and is thus similar to social exchange theory (e.g., Homans, 1961; Thibaut & Kelly, 1959). The time dimension emphasized by social penetration theory is not crucial in social exchange formulations, and social penetration theory is unique in its emphasis on the joint effects of verbal and nonverbal communication in the acquaintance process.

---

[10] L. P. Frankfurt. The role of some individual and interpersonal factors on the acquaintance process. Unpublished doctoral dissertation, American University, 1965.

[11] C. H. Swensen. Self-disclosure as a function of personality variables. Paper presented at the meeting of the American Psychological Association, San Francisco, September 1968.

Reciprocity of Self-Disclosure

Jourard (1959b) measured self-disclosure within a group of nine female nursing college faculty members. The amount disclosed to a given colleague correlated highly with the amount of disclosure received from that colleague. Jourard and Landsman (1960) replicated the above finding in a group of nine male graduate students. Further evidence for the reciprocity effect was obtained by Jourard and Richman (1963) who correlated subjects' reports of disclosure output and input (disclosures received) for mother, father, best male friend, and best female friend.

Other researchers have extended the investigation of reciprocity to different relationships and types of subjects. Levinger and Senn (1967) studied reciprocity of disclosure of feelings in marital relationships. The correlation between reported output and reported input was .91 for husbands and .79 for wives. Levinger and Senn point out that (like the Jourard and Richman, 1963, study) the input and output scores were obtained from the same person, and are probably artificially high. Correlations between husband output or input and wife output or input were considerably lower, averaging about .50.

Skypeck[12] and Rivenbark (1971) report high correlations between reported output and input in samples of subjects ranging from 6 to 18 years of age.

Several experiments have confronted subjects with a confederate who discloses at either high or low levels. Regardless of whether disclosure was manipulated by varying intimacy (Ehrlich & Graeven, 1971; Murdoch, Chenowith, & Rissman[13]), or number of statements disclosed (Chittick & Himelstein, 1967), these studies have found that the high-disclosing confederate elicits greater self-disclosure from subjects than the low-disclosing confederate.

One potential criticism of these studies is that the subject may be imitating the response of the confederate rather than engaging in a reciprocation process. However, Tognoli (1969) did not allow subjects to imitate by eliminating the confederate's statement as an option for the subject's response, and still obtained evidence for reciprocity of disclosure.

Levin and Gergen (1969) suggested that medium amounts of disclosure from another person indicate his desire for a closer relationship and his trustfulness, yet another who communicates a great deal about himself may be seen to lack discretion and to be untrustworthy, and predicted a curvilinear relationship between other disclosure and subject disclosure. Subjects received a 40-item self-rating form from a partner on which either 4, 16, or 32 items were checked. Rather than finding a curvilinear relationship, subjects revealed a mean total of 10.5, 17.2, and 22.0 items in the low, moderate, and high groups, respectively. Levin and Gergen (1969) concluded that "the more information revealed by the other, the greater the absolute amount returned. On the other hand, in economic terms, the subjects reveal proportionately less as the partner reveals more and more [p. 448]."

---

[12] G. Skypeck. Self-disclosure in children, ages six through twelve. Unpublished master's thesis, University of Florida, 1967.

[13] P. Murdoch, R. Chenowith, & K. Rissman. Eligibility and intimacy effects on self-disclosure. Paper presented at the meeting of the Society of Experimental Social Psychology, Madison, Wisconsin, October 31–Nov. 1, 1969.

Worthy, Gary, and Kahn (1969) suggested that reception of self-disclosing information from another is rewarding since reception of such information implies that one is trusted, and that more intimate information represents greater reward for the receiver. Worthy et al. tested several predictions derived from social exchange theory, among them the principle that outcomes exchanged will be of comparable value. Female subjects were tested in groups of four. On each of 10 trials, each subject chose questions to answer and send to each of the three other subjects. The questions had been prescaled for intimacy value. Support was obtained for the reciprocity hypothesis: The mean levels of disclosure (on a 7-point scale) to the other who had disclosed least, next, and most were 2.997, 3.316, and 4.041, respectively.

Cozby (1972) attempted to extend the social exchange theory analysis of Worthy et al. (1969). Cozby argued that, while increasing intimacy of disclosure may represent increasing reward, various costs become salient at high intimacy levels, the most obvious being anxiety over revealing information one would rather keep private, and having to interact with an individual who is behaving in a somewhat unusual manner. This analysis is supported by Argyle and Kendon's (1967) description of increasing intimacy in interpersonal relationships until halted by anxiety. Subjects were shown 10 topics which a hypothetical other had disclosed. These were low, medium, or high in intimacy value. They then chose 10 out of 70 items to disclose to the other. It was predicted that reciprocity would determine subjects' responses to the low- and medium-disclosing other, but that intimacy level would not increase (or possibly decrease) as other disclosure increased from medium to high intimacy. The results represent an independent replication of the Levin and Gergen (1969) experiment: A small (but significant at the .05 level) increase was observed as other disclosure increased from medium to high intimacy.

Reciprocity appears to become less powerful as a determinant of subjects' responses at high levels of intimacy, but there is as yet no explanation for the counterintuitive results obtained by Levin and Gergen, and Cozby. It is possible that making self-disclosures is rewarding for the person who is disclosing, and the laboratory setting is a particularly likely place for this to occur. In other words, the reward effect may tend to outweigh the anxiety discussed previously in a laboratory setting in which subjects do not see each other, and there is little likelihood of future interactions. The Murdoch et al. (see Footnote 13) experiment is relevant to this discussion. Murdoch et al. varied whether or not there was a possibility of future interaction with the confederate. Subjects disclosed more to the confederate with whom there was no possibility of future interactions. It is possible that the curvilinear effect would be obtained in a situation in which there was a possibility of future interaction with the other, or possibly if subject anonymity was precluded by previous encounter between the subject and confederate.

Self-Disclosure and Liking

Jourard and Lasakow (1958) found that disclosure to mother and father correlated significantly with liking ($r = .63$ and .53, respectively). In a subsequent investigation, Jourard (1959b) had each of nine female nursing college faculty members order the other eight subjects in terms of liking, and found disclosure positively related to liking; Jourard and Landsman (1960) replicated the Jourard (1959b) procedure with a group of nine male graduate students. The relationship between disclosure and liking was not obtained with this male sample. Jourard and Landsman (1960) concluded that their results, along with the findings of Jourard (1959b), corroborate the "statements abounding in the popular literature which portray men as strong and silent and distrustful of feelings; women are commonly portrayed as more emotional, talkative, overly trusting of those whom they like, and hence more vulnerable to betrayal than men [p. 185]." Alternatively, it is possible that the male group was a much less cohesive group, with fewer interaction experiences across the 36 relationships than the nursing faculty group.

Other questionnaire studies finding a relationship between disclosure and liking include Halverson and Shore (1969), Fitzgerald (1963), and Altman and Haythorn (1965).

Frankfurt (see Footnote 10) had female subjects indicate what they would disclose to another girl described as being liked or disliked in a series of scenarios which represented successive stages in the relationship. More was disclosed to the liked than disliked girl, and the effect occurred only in intimate topic areas.

Worthy et al. (1969) predicted that liking leads to disclosure to the other, and also that disclosure from another will lead to greater liking. Subjects, in groups of four, had a 10-minute "get acquainted" session, and then indicated their liking for the other subjects. In the actual experiment, each subject chose items of information to disclose to each other subject on each of 10 trials. On the first trial, it was found that most was disclosed to the most liked other, and least was disclosed to the least liked other. In addition, analysis of covariance of final liking scores with initial liking covaried showed that subjects liked most those others who had disclosed most to them.

Cozby (1972) suggested that the relationship between self-disclosure and liking is curvilinear. The high-disclosing other may arouse anxiety with his overly intimate disclosures, and so the high discloser may be seen as providing the subject with negative outcomes. Strong support for the predicted curvilinear relationship was obtained in a situation in which subjects were responding to a hypothetical other.

The Worthy et al. and Cozby studies used female subjects. A study by Ehrlich and Graeven (1971) used male subjects who conversed for 16 minutes with a male confederate who disclosed at either low- or high-intimacy values. There were no differences in rated liking of the confederate in the two conditions. This finding may support Jourard and Landsman's (1960) suggestion that the relationship between disclosure and liking may be different for males and females. However, interpretation of liking ratings is made difficult by a heretofore unmentioned aspect of these disclosure studies—content. In all the studies except Cozby (1972) which used a role-playing procedure, subjects received from another content about a topic which varied in intimacy value. To illustrate, the topic, "favorite television program," is low in intimacy value. In an actual interaction, this topic becomes a statement such as "My favorite television program is Mod Squad." The problem becomes obvious: Is the subject's liking rating due to the intimacy value of the topic or his own feelings about that particular program? The variables of topic intimacy and similarity–dissimilarity of the content should be combined factorially. From the Cozby (1972) study and the work of Byrne (1969), it might be expected that the most liked other would be one who disclosed topics of medium intimacy and expressed views that were similar to those of the subject.

Pederson and Higbee (1969b) noted that the Jourard and Landsman (1960) findings might indicate that the characteristics of the target person that determine disclosure may vary from one target person to another. Pederson and Higbee correlated disclosure scores for mother, father, best opposite-sex friend and best same-sex friend with ratings of these targets on 11 adjectives. For both males and females, liking was significantly related to disclosure only for father and opposite-sex friend. For each target, there were sex differences in the adjectives that were related to self-disclosure. Interestingly, for both sexes, few characteristics of the same sex friend were related to disclosure to the friend.

Rubin (1970) has recently provided distinctions between liking and loving relationships. Although Rubin's (1970) love and liking scales have not been correlated with self-disclosure, there are indications that the distinction may prove useful in the study of self-disclosure. For instance, Rubin (1970) found that females love (but not like) their same-sex friends more than males do, which provides a reason for the finding that women tend to disclose to their same-sex friends more than males do (Jourard & Lasakow, 1958).

The finding that disclosure to spouse is greater than to any other target (Jourard & Lasakow, 1958) suggests that love results in greater disclosure than liking. Within the marital relationship itself, it might be expected that self-disclosure would be related to marital satisfaction. However, empirical support for such a relationship is weak. Katz et al. (1963) report that, for males, marital satisfaction is related to greater disclosures of anxieties and worries, but not of other topics. No relationship was found for wives. Levinger and Senn (1967) found that, in general, disclosure tends to be related to satisfaction, but that the correlation is higher between satisfaction and disclosure of pleasant feelings, contrary to the expectation that a strong marital bond would be characterized by the ability to openly discuss unpleasant feelings. However, this finding would be expected if we assume that the disclosure of pleasant feelings is equivalent to exchanging positive rewards. Shapiro and Swensen (1969) found no relationship between marital satisfaction and disclosure.

Social Approval

In an experiment by Colson[14] a confederate responded to subjects' disclosures with positive, neutral, or negative evaluations. Disclosure was greatest in the positive evaluation condition and lowest in the negative evaluation condition.

In an experiment by Taylor et al. (1969), a confederate responded to subjects' disclosures in four 45-minute interaction sessions according to one of four reinforcement schedules: (*a*) continuous positive, (*b*) early negative, later positive, (*c*) continuous negative, or (*d*) early positive, later negative. As predicted, subjects in the continuous and later positive reinforcement conditions disclosed more about themselves, talked longer, and were generally more intimate. For duration and amount of disclosure, the differences were primarily at intimate areas of exchange. No contrast effect for disclosure or liking of the confederate was found in the later positive and later negative groups.

Dependency and Self-Disclosure

Altman and Haythorn (1965) investigated the self-disclosure patterns of pairs of naval recruits who were isolated in a small room for 10 days (high mutual dependency). A control condition consisted of pairs who spent most of their waking hours together, but had access to the outside world (low mutual dependency). The high-dependency dyads were more intimate and showed a more active pattern of social interaction than did the low-dependency dyads.

Thibaut and Kelly (1959) suggested that the formality and constraint which characterizes first encounters functions to prevent forming relationships that may prove unsatisfying in the future. A total lack of dependency, as with interaction with a stranger with whom there is little likelihood of future interaction, removes the constraints and causes increased disclosure. Experimental support for the "stranger on a train" phenomenon had been demonstrated by Murdoch et al. (see Footnote 13). Taylor et al. (1969) obtained no effect of a 3 week versus 6 month interaction expectancy, suggesting that greater disclosure to a "stranger" may occur only when the expected interaction period is very brief.

There is also evidence that individuals will resist entering into relationships in which the other person holds a great deal of power. Kounin, Polansky, Biddle, Coburn, and Fenn (1956) found that persons feel more at ease and reveal more negative things about themselves with a nonpowerful counselor. However, under some circumstances, people may disclose more to a person with high power than to a person with low power. Slobin, Miller, and Porter (1968) found that, while greatest disclosure is made to fellow workers in a business organization, there is greater disclosure to immediate superiors than to immediate subordinates. Perhaps disclosure to superiors is an ingratiation technique.

Disclosure as a Function of Personality of the Interactants

Social penetration theory asserts that amount of mutual disclosure in a dyad is a function of the particular personality characteristics of the two individuals. The obvious prediction that pairs of high disclosers will disclose more to one another than pairs of low disclosers has been confirmed in several studies (Jourard & Resnick, 1970; Taylor, 1968). Jourard and Resnick also found that when a low-disclosing subject is paired with a high discloser, the subject increases his disclosure output to match the level of the high discloser.

---

[14] W. N. Colson. Self-disclosure as a function of social approval. Unpublished master's thesis, Howard University, 1968.

Pairs of subjects in a study by Altman and Haythorn (1965) were either homogeneously high, heterogeneous, or homogeneously low on need achievement, need affiliation, need dominance, and dogmatism. Although there were no significant effects of group composition on self-disclosure, perhaps due to small *N*, high need achievers disclosed more than low achievers, and low-dominance subjects disclosed more than high-dominance subjects in intimate topic areas while the opposite was true in nonintimate topic areas.

Swensen and Nelson[15] studied disclosure in dyads matched for similarity and dissimilarity in extraversion, neuroticism, and attitudes. Disclosure was highest in male pairs who were similar on extraversion but different on neuroticism, in female pairs who were different on attitudes, and in male–female pairs who were similar on all three variables. Persons and Marks (1970) found that greatest intimacy occurred in an interview situation when both interviewer and interviewee has the same MMPI code type (4-2, 4-8, or 4-9) than when they were paired in nonmatched combinations.

Self-Disclosure over Time

Taylor (1968) administered a self-disclosure questionnaire to male freshmen roommates after they had known each other for 1, 3, 6, 9, and 13 weeks. Half of the roommate pairs were both high revealers, while the other half were both low revealers. At all points in time, the high-revealing dyads reported more mutual disclosure than did the low-revealing dyads, although the rate of the increase over time was approximately the same for both groups. There was a rapid increase in nonintimate disclosures, and a slow, gradual increase in intimate disclosures over time. Measures of liking showed no significant differences between the low- and high-revealing dyads. However, both groups showed a significant decrement in liking over time, and this trend was more pronounced among the high revealers. This decrement in liking with a concomitant increase in disclosure may mean that "familiarity breeds contempt," but it may also indicate that most randomly chosen dyads are not compatible enough to be comfortable at prolonged high disclosure.

Frankfurt (see Footnote 10) obtained evidence of increasing amount and intimacy of disclosure as subjects described what they would disclose to another person at successive stages in their relationship. A large increase in number of intimate topics discussed was found for subjects who liked the other person, while subjects who did not like the other person showed only a slight increase.

In the Taylor et al. (1969) study, favorable reward/cost groups showed a greater increase in duration of time talked over four 45-minute interaction periods than did negative reward/cost groups, and the effect occurred primarily in intimate areas of exchange. Contrary to expectations, greatest intimacy occurred during the first period of interaction, although a gradual increase from second to fourth period was found for all groups combined.

Methodological problems make research on the developmental aspects of interpersonal relationships especially difficult. Observation of behavior over a long period of time in the laboratory, as in the Taylor et al. (1969) study, may provide insights into the developmental process, but it seems inevitable that factors such as fatigue will affect the results. An alternative approach might be to study people who have known each other for varying lengths of time. For instance, Jourard (1961a) administered the JSDQ to college students ranging in age from 17 to 55 years. It was found that disclosure to parents decreased with age, while disclosure to opposite-sex friend or spouse increased up to age 40, after which a decrease was observed. A problem with the Jourard study is that not all subjects were married, and the effect of length of marriage is confounded with age.

---

[15] C. H. Swensen, & D. A. Nelson. Interpersonal attraction as a function of personality and attitudes. Paper presented at the meeting of the Midwestern Psychological Association, Chicago 1967.

but the study does illustrate the type of approach that might be taken.

DISCLOSURE IN THERAPY AND EXPERIMENTS

While the research reviewed in this section could have been placed within the context of "interpersonal relationships," the experimenter–subject or therapist–client relationship would seem to merit special consideration. The therapist, interviewer, and experimenter, whether it be for the person's own benefit or for scientific reasons, seeks information about another person's private self. The studies reviewed in this section indicate that the effects of therapist or experimenter disclosure may have a number of practical and methodological implications for therapy outcome and experimental results.

Several authors (Jourard, 1964, 1968; Mowrer, 1964; Rogers, 1961) have written on the importance of full client disclosure for successful therapy. Truax and Carkhuff (1965) have reported significant correlations between therapist and patient disclosure. Further, level of patient disclosure appears to be a predictor of final case outcome (Truax & Carkhuff, 1965).

Powell (1968) found that subjects disclosed more when the interviewer responded to subjects' self-references with open disclosure than when he used approval–supportive or reflection–restatement techniques. However, Vondracek (1969) reported greater amounts of disclosure when the interviewer used a "probing" technique than when he was "reflecting" or "revealing." It is possible that "probing" is more effective than self-revelation in eliciting highly intimate information (Tuckman, 1966).

Jourard and Jaffee (1970) found that increases in the duration of interviewer disclosure prior to discussing a topic resulted in increases in duration of subject disclosure on the topic.

Vondracek and Vondracek (1971) have shown that sixth-grade children disclose more to an adult interviewer who discloses than to an interviewer who does not disclose. There was also evidence that the increased disclosure was in the same content area disclosed by the interviewer. Ehrlich and Graeven (1971) did not find evidence of topical reciprocity in their male college student sample. Future research should determine the conditions under which topical reciprocity will occur.

The interviewer or experimenter who discloses, in addition to eliciting greater disclosure from subjects (i.e., inducing a reciprocity effect), is rated as more trustworthy (Drag, see Footnote 3) and more positively in general (Jourard & Friedman, 1970) than the experimenter who does not self-disclose.

Objections to the use of therapist disclosure to increase patient disclosure have been raised by Vondracek and Vondracek (1971) and Polansky (1967) who states that the technique is clinically very sloppy. Although these authors do not elaborate on their objections, perhaps they feel that disclosure to certain patients would have an adverse effect on the course of therapy, or that certain therapists would feel uncomfortable disclosing and communicate the discomfort to the patient.

Disclosure may also affect the nature of subjects' responses in experiments. In general, studies which manipulate experimenter disclosure involve a period of experimenter–subject interaction of approximately 30 minutes in duration in which the experimenter and subject discuss their "personal backgrounds and interests, problems, and hopes and aspirations [Jourard, 1969, p. 113]." Control subjects have no prior contact with the experimenter. Subjects who receive experimenter disclosure have been shown to change their Edwards Personal Preference Schedule responses (Jourard & Kormann, 1968), give more revealing responses on the Rotter Incomplete Sentence Blank (Heifitz[16]), and require fewer trials to criterion on a paired-associate learning task (Frey[17]).

---

[16] M. L. Heifitz. Experimenter effect upon openness of response to the Rotter Incomplete Sentence Blank. Unpublished honors paper, University of Florida, 1967.

[17] M. Frey. The effects of self-disclosure and social reinforcement on performance in paired-associate learning. Unpublished honors paper, University of Florida, 1968.

Drag (1969) studies the effect of experimenter participation in discussion groups composed of two, four, or eight subjects whose task was to disclose themselves on various topics. In groups in which the experimenter did not disclose, Drag (1969) replicated the common finding that smaller groups provide the participants more opportunity to interact or express themselves (Thomas & Fink, 1963). When the experimenter joined in the discussion and self-disclosed, amount of subject disclosure was not affected in the two- or eight-person groups. However, in the four-person group, subject disclosure was similar to disclosure scores in the two-person group, and significantly greater than in the four-person–no-experimenter-disclosure group. This result was not expected, for it had been predicted that amount of disclosure would increase by equal amounts in all three groups when the experimenter disclosed. The finding would seem to indicate that in the eight-person group, experimenter disclosure (or probably disclosure by any group member) is not sufficient to overcome the inhibiting effects of the large group.

Jourard (1969) has argued that disclosure by experimenters will result in greater honesty by subjects, and also prevent experimenters from acting like "spies" and inhuman manipulators. An equity theory (Adams, 1965) analysis would suggest that the experimenter who asks subjects to fill out questionnaires, yet who tells nothing of himself, may have created a situation of inequity. If this is actually the case, subjects may attempt to restore equity by lying, or being careless in their responses (see Weick, 1970).

The equity theory analysis rests upon the assumption that subjects perceive inequity in the situation. It is possible, however, that the subject actually expects to disclose in the experimental situation, and that the opportunity to disclose is a major reason for coming to the experiment (Hood & Back, 1971). Hood and Back found that, among males, volunteers for social psychological experiments are higher disclosers than nonvolunteers. Among females, JSDQ scores were related to volunteering only for experiments which seemed to indicate that subjects would be required to make statements about themselves.

Future research should determine the conditions under which experimenter or subject disclosure will affect experimental results, and the nature of the effects themselves. One possibility that deserves investigation is whether disclosure by the experimenter has the effect of making subjects more sensitive to demand characteristics (Orne, 1962) or experimenter biases (Rosenthal, 1966). If so, we can imagine subjects giving dishonest responses in order to try to please the experimenter and confirm his hypotheses. Future research in this area should also include controls for amount of time the experimenter spends with the subject (i.e., a separate condition with very nonintimate disclosure).

SELF-DISCLOSURE AND OTHER "INTIMACY" VARIABLES

The affiliative conflict theory (Argyle & Dean, 1965) asserts that an equilibrium for "intimacy" exists for any pair of individuals, and that the equilibrium point is a joint function of eye contact, physical proximity, amount of smiling, intimacy of topic (i.e., self-disclosure), and other unspecified variables. The theory states that if the equilibrium is disturbed along any one of its dimensions, attempts will be made to restore it by adjusting the others. A study by Exline, Gray, and Schutte (1965) found support for the theory in an experiment which related eye-contact to intimacy of interview material. Subjects spent significantly less time looking at the interviewer in an intimate interview condition than in a nonintimate interview.

Jourard and Friedman (1970) examined the relationship between experimenter–subject "distance" and self-disclosure. In Experiment 1, females (but not males) increased duration of disclosure as distance increased, thus providing partial support for the theory. Distance in this study was varied by using a combination of eye contact and physical distance. In Experiment 2, distance was varied by a combination of the experimenter touching and/or disclosing to the subject. Regardless of sex, subjects *increased* disclosure time as distance *decreased*. The authors concluded that the Argyle and Dean equilibrium phenomenon will not be obtained in situations in which there is trust and positive feelings toward the person who comes close. However, the use of self-disclosure as both independent and dependent variable neglects the multivariate nature of the Argyle and Dean theory, and leads to the suggestion that the results of Jourard and Friedman's second experiment do little more than provide additional confirmation of the reciprocity rule in self-disclosure.

Regardless of the validity of the Argyle and Dean (1965) theory, it is obvious that people do engage in nonverbal communication, vary their speaking distance, and use objects in the physical environment while communicating verbally. Altman and Taylor (1973) point out that research on nonverbal, environmental, and verbal behaviors has proceeded in parallel fashion, but that no attempt has been made to integrate the behaviors into a rich description of interpersonal relationships.

CONCLUSION

Simmel (1964) writes that "obviously, all relations which people have to one another are based on their knowing something about one another [p. 307]." This statement seems so intuitively obvious that it is surprising that there is not more research or theoretical development in the area of self-disclosure. It also seems obvious that there are individual differences in self-disclosure, yet we know little about the meaning of these differences.

Self-disclosing behavior may be seen as the product of two opposing forces, one operating to increase disclosure, the other operating to inhibit disclosure. The first force is the one studied most extensively by disclosure researchers. There are also factors which operate to inhibit disclosure. These might be termed discretion, or a need for privacy, and have been neglected in research on self-disclosure. Rather, self-disclosure has been closely associated with the encounter group movement which emphasizes the need to relate, communicate, and be honest and open with others. Self-disclosure has thus been made a positive value, something to be fostered. This orientation to self-disclosure has been expressed by Bennett (1967):

> The critical problem we face is not how to keep secrets from each other, but how to facilitate this readiness to communicate. The overriding question is how to maintain an atmosphere of trust and confidence which will enable us to talk about personal affairs as freely as we talk about automobiles; to share experience as we share the weather [p. 375].

What happens to privacy when we share our intimate thoughts as easily as we talk about the weather? Bennett classifies people who would withhold information as hermits, prudes, paranoids, or rascals. Yet Altman and Taylor (1973) have expressed fear of a "tyranny of openness" in which individuals have no freedom to have private thoughts because full disclosure is demanded by others in all social encounters.

Simmel (1964) has written extensively on the importance of discretion, and suggests that "an ideal sphere lies around every human being . . . this sphere cannot be penetrated, unless the personality value of the individual is thereby destroyed [p. 321]." Thus, self-concealment (not lying or misrepresenting one's self, but maintaining a private area of the self) may have functional significance in giving a person a sense of "individuality." Simmel (1964) also suggests that some marital difficulties are the result of too much self-disclosure: "They lapse into a matter-of-factness which has no longer any room for surprises [p. 329]." Thus, discretion may be important in preventing boredom from occurring in interpersonal relationships.

It is hoped that the peculiar value orientation that has accompanied much of the written discussion of self-disclosure can be dispensed with. It seems far more important that the research be conducted. We should know more about the relationship between verbal and nonverbal components of messages (Duncan, 1969). We can ask about the nature of self-disclosure when people withdraw from a relationship (Altman & Taylor, 1973). What is the function of privacy in interpersonal relationships? What are the consequences of unrestrained self-disclosure? So much remains to be learned about this important aspect of interpersonal relationships that value-laden speculations that close off areas of investigation can only be seen as a hinderance to our understanding of self-disclosure.

REFERENCES

Adams, J. S. Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 2. New York: Academic Press, 1965.

Altman, I., & Haythorn, W. W. Interpersonal exchange in isolation. *Sociometry,* 1965, **28**, 411–426.

Altman, I., & Taylor, D. A. *Social penetration: The development of interpersonal relationships*. New York: Holt, Rinehart & Winston, 1973, in press.

Argyle, M., & Dean, J. Eye-contact, distance, and affiliation. *Sociometry,* 1965, **28**, 289–304.

Argyle, M., & Kendon, A. The experimental analysis of social performance. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 3. New York: Academic Press, 1967.

Axtell, B., & Cole, C. W. Repression-sensitization response mode and verbal avoidance. *Journal of Personality and Social Psychology,* 1971, **18**, 133–137.

Bennett, C. C. What price privacy? *American Psychologist,* 1967, **22**, 371–376.

Bennett, E. M., & Cohen, L. R. Men and women: Personality patterns and contrasts. *Genetic Psychology Monographs,* 1959, **59**, 101–155.

Brown, R. *Social psychology*. New York: Free Press, 1965.

Byrne, D. Attitudes and attraction. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 4. New York: Academic Press, 1969.

Burhenne, D., & Mirels, H. L. Self-disclosure in self-descriptive essays. *Journal of Consulting and Clinical Psychology,* 1970, **35**, 409–413.

Campbell, D. T., & Fiske, D. W. Convergent and discriminating validation by the multitrait-multimethod matrix. *Psychological Bulletin,* 1959, **56**, 81–105.

Chittick, E. V., & Himelstein, P. The manipulation of self-disclosure. *Journal of Psychology,* 1967, **65**, 117–121.

Cozby, P. C. Self-disclosure, reciprocity, and liking. *Sociometry,* 1972, **35**, 151–160.

Dimond, R. E., & Hellkamp, D. T. Race, sex, ordinal position or birth, and self-disclosure in high school students. *Psychological Reports,* 1969, **25**, 235–238.

Dimond, R. E., & Munz, D. C. Ordinal position of birth and self-disclosure in high school students. *Psychological Reports,* 1967, **21**, 829–833.

Doster, J. A., & Strickland, B. R. Perceived child-rearing practices and self-disclosure patterns. *Journal of Consulting and Clinical Psychology,* 1969, **33**, 382.

Drag, R. M. Self-disclosure as a function of group size and experimenter behavior. *Dissertation Abstracts International,* 1969, **30**(5-B), 2416.

Duncan, S., Jr. Nonverbal communication. *Psychological Bulletin,* 1969, **72**, 118–137.

Ehrlich, H. J., & Graeven, D. B. Reciprocal self-disclosure in a dyad. *Journal of Experimental Social Psychology,* 1971, **7**, 389–400.

Exline, R., Gray, D., & Schutte, D. Visual behavior in a dyad as affected by interview content and sex of respondent. *Journal of Personality and Social Psychology,* 1965, **1**, 201–209.

Fitzgerald, M. P. Self-disclosure and expressed self-esteem, social distance, and areas of the self revealed. *Journal of Psychology,* 1963, **56**, 405–412.

Halverson, C. F., Jr., & Shore, R. E. Self-disclosure and interpersonal functioning. *Journal of Consulting and Clinical Psychology,* 1969, **33**, 213–217.

Himelstein, P., & Lubin, B. Attempted validation of the Self-Disclosure Inventory by the peer nomination technique. *Journal of Psychology,* 1965, **61**, 13–16.

Himelstein, P., & Kimbrough, W. W., Jr. A study of self-disclosure in the classroom. *Journal of Psychology,* 1963, **55**, 437–440.

Homans, G. C. *Social behavior: Its elementary forms*. New York: Harcourt, Brace, & World, 1961.

Hood, T. C., & Back, K. W. Self-disclosure and the volunteer: A source of bias in laboratory experiments. *Journal of Personality and Social Psychology,* 1971, **17**, 130–136.

Hurley, J. R., & Hurley, S. J. Toward authenticity in measuring self-disclosure. *Journal of Counseling Psychology,* 1969, **16**, 271–274.

Jaffee, L. D., & Polansky, N. A. Verbal inaccessibility in young adolescents showing delinquent trends. *Journal of Health and Human Behavior,* 1962, **3**, 105–111.

Jennings, F. L. Religious beliefs and self-disclosure. *Psychological Reports,* 1971, **28**, 193–194.

Jourard, S. M. Healthy personality and self-disclosure. *Mental Hygiene,* 1959, **43**, 499–507. (a)

Jourard, S. M. Self-disclosure and other-cathexis. *Journal of Abnormal and Social Psychology,* 1959, **59**, 428–431. (b)

Jourard, S. M. Age trends in self-disclosure. *Merrill-Palmer Quarterly,* 1961, **7**, 191–197. (a)

Jourard, S. M. Religious denomination and self-disclosure. *Psychological Reports,* 1961, **8**, 446. (b)

Jourard, S. M. Rorschach productivity and self-disclosure. *Perceptual and Motor Skills,* 1961, **13**, 232. (c)

Jourard, S. M. Self-disclosure scores and grades in nursing college. *Journal of Applied Psychology,* 1961, **45**, 244–247. (d)

Jourard, S. M. *Personal adjustment: An approach through the study of healthy personality*. (2nd ed.) New York: Macmillan, 1963.

Jourard, S. M. The transparent self. Princeton: Van Nostrand, 1964.

Jourard, S. M. Disclosing man to himself. Princeton: Van Nostrand, 1968.

Jourard, S. M. The effects of experimenters' self-disclosure on subjects' behavior. In C. Spielberger

(Ed.), *Current topics in community and clinical psychology*. New York: Academic Press, 1969.

Jourard, S. M., & Friedman, R. Experimenter-subject "distance" and self-disclosure. *Journal of Personality and Social Psychology,* 1970, **15**, 278–282.

Jourard, S. M., & Jaffee, P. E. Influence of an interviewer's disclosure on the self-disclosing behavior of interviewees. *Journal of Counseling Psychology,* 1970, **17**, 252–257.

Jourard, S. M., & Kormann, L. A. Getting to know the experimenter, and its effect on psychological test performance. *Journal of Humanistic Psychology,* 1968, **8**, 155–159.

Jourard, S. M., & Landsman, M. J. Cognition, cathexis, and the "dyadic effect" in men's self-disclosing behavior. *Merrill-Palmer Quarterly,* 1960, **6**, 178–186.

Jourard, S. M., & Lasakow, P. Some factors in self-disclosure. *Journal of Abnormal and Social Psychology,* 1958, **56**, 91–98.

Jourard, S. M., & Resnick, J. L. The effect of high revealing subjects on the self-disclosure of low revealing subjects. *Journal of Humanistic Psychology,* 1970, **10**, 84–93.

Jourard, S. M., & Richman, P. Disclosure output and input in college students. *Merrill-Palmer Quarterly,* 1963, **9**, 141–148.

Kaplan, M. F. Interview interaction of repressors and sensitizers. *Journal of Consulting Psychology,* 1967, **31**, 513–516.

Katz, I., Goldston, J., Cohen, M., & Stucker, S. Need satisfaction, perception, and cooperative interaction in married couples. *Marriage and Family Living,* 1963, **25**, 209–214.

Kounin, J., Polansky, N., Biddle, B., Coburn, H., & Fenn, A. Experimental studies of clients' reactions to initial interviews. *Human Relations,* 1956, **9**, 265–293.

Levin, F. M., & Gergen, K. J. Revealingness, ingratiation, and the disclosure of self. *Proceedings of the 77th Annual Convention of the American Psychological Association,* 1969, **4**(Pt. 1), 447–448.

Levinger, G., & Senn, D. J. Disclosure of feelings in marriage. *Merrill-Palmer Quarterly,* 1967, **13**, 237–249.

Lewin, K. Some social-psychological differences between the United States and Germany. In G. Lewin (Ed.), *Resolving social conflicts: Selected papers on group dynamics, 1935–1946*. New York: Harper, 1948.

Lubin, B., & Harrison, R. L. Predicting small group behavior with the Self-Disclosure Inventory. *Psychological Reports,* 1964, **15**, 77–78.

Maslow, A. H. *Motivation and personality*. New York: Harper, 1954.

Mayer, J. E. Disclosing marital problems. *Social Casework,* 1967, **48**, 342–351.

Mayo, P. R. Self-disclosure and neurosis. *British Journal of Social and Clinical Psychology,* 1968, **7**, 140–148.

Melikian, L. H. Self-disclosure among university students in the Middle East. *Journal of Social Psychology,* 1962, **57**, 257–263.

Mowrer, O. H. *The new group therapy*. Princeton: Van Nostrand, 1964.

Orne, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist,* 1962, **17**, 776–783.

Pederson, D. M., & Breglio, V. J. Personality correlates of actual self-disclosure. *Psychological Reports,* 1968, **22**, 495–501. (a)

Pederson, D. M., & Breglio, V. J. The correlation of two self-disclosure inventories with actual self-disclosure: A validity study. *Journal of Psychology,* 1968, **68**, 291–298. (b)

Pederson, D. M., & Higbee, K. L. An evaluation of the equivalence and construct validity of various measures of self-disclosure. *Educational and Psychological Measurement,* 1968, **28**, 511–523.

Pederson, D. M., & Higbee, K. L. Personality correlates of self-disclosure. *Journal of Social Psychology,* 1969, **78**, 81–89. (a)

Pederson, D. M., & Higbee, K. L. Self-disclosure and relationship to the target person. *Merrill-Palmer Quarterly,* 1969, **15**, 213–220. (b)

Persons, R. W., & Marks, P. A. Self-disclosure with recidivists: Optimum interviewer-interviewee matching. *Journal of Abnormal Psychology,* 1970, **76**, 387–391.

Plog, S. C. The disclosure of self in the United States and Germany. *Journal of Social Psychology,* 1965, **65**, 193–203.

Polansky, N. A. The concept of verbal accessibility. *Smith College Studies in Social Work,* 1965, **36**, 1–46.

Polansky, N. A. On duplicity in the interview. *American Journal of Orthopsychiatry,* 1967, **37**, 568–580.

Polansky, N. A., & Brown, S. Q. Verbal accessibility and fusion fantasy in a mountain county. *American Journal of Orthopsychiatry,* 1967, **37**, 651–660.

Powell, W. J. Differential effectiveness of interviewer interventions in an experimenter interview. *Journal of Consulting and Clinical Psychology,* 1968, **32**, 210–215.

Powell, W. J., & Jourard, S. M. Some objective evidence of immaturity in under-achieving college students. *Journal of Counseling Psychology,* 1963, **10**, 276–282.

Rickers-Ovsiankina, M. A. Social accessibility in three age groups. *Psychological Reports,* 1956, **2**, 283–294.

Rickers-Ovsiankina, M. A., & Kusmin, A. A. Individual differences in social accessibility. *Psychological Reports,* 1958, **4**, 391–406.

Rivenbark, W. H., III. Self-disclosure patterns among adolescents. *Psychological Reports,* 1971, **28**, 35–42.

Rogers, C. R. *On becoming a person*. Boston: Houghton Mifflin, 1961.

Rosenthal, R. *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts, 1966.

RUBIN, Z. Measurement of romantic love. *Journal of Personality and Social Psychology*, 1970, **16**, 265–273.

SHAPIRO, A., & SWENSEN, C. Patterns of self-disclosure among married couples. *Journal of Counseling Psychology*, 1969, **16**, 179–180.

SIMMEL, G. The secret and the secret society. In K. Wolff (Ed.), *The sociology of Georg Simmel*. New York: Free Press, 1964.

SLOBIN, D. I., MILLER, S. H., & PORTER, L. W. Forms of address and social relations in a business organization. *Journal of Personality and Social Psychology*, 1968, **8**, 289–293.

SMITH, M. B. Research strategies toward a conception of positive mental health. *American Psychologist*, 1959, **14**, 673–681.

STANLEY, G., & BOWNES, A. F. Self-disclosure and neuroticism. *Psychological Reports*, 1966, **18**, 350.

TAYLOR, D. A. The development of interpersonal relationships: Social penetration processes. *Journal of Social Psychology*, 1968, **75**, 79–90.

TAYLOR, D. A., & ALTMAN, I. Intimacy-scaled stimuli for use in studies of interpersonal relations. *Psychological Reports*, 1966, **19**, 729–730.

TAYLOR, D. A., ALTMAN, I., & SORRENTINO, R. Interpersonal exchange as a function of rewards and costs and situational factors: Expectancy confirmation-disconfirmation. *Journal of Experimental Social Psychology*, 1969, **5**, 324–339.

TAYLOR, D. A., & OBERLANDER, L. Person-perception and self-disclosure: Motivational mechanisms in interpersonal processes. *Journal of Experimental Research in Personality*, 1969, **4**, 14–28.

THIBAUT, J. W., & KELLY, H. H. *The social psychology of groups*. New York: Wiley, 1959.

THOMAS, E. J., & FINK, C. F. Effects of group size. *Psychological Bulletin*, 1963, **60**, 371–384.

TOGNOLI, J. Response matching in interpersonal information exchange. *British Journal of Social and Clinical Psychology*, 1969, **8**, 116–123.

TRUAX, C. B., & CARKHUFF, R. R. Client and therapist transparency in the psychotherapeutic encounter. *Journal of Counseling Psychology*, 1965, **12**, 3–9.

TRUAX, C., & WITTMER, J. Self-disclosure and personality adjustment. *Journal of Clinical Psychology*, in press.

TUCKMAN, B. W. Interpersonal probing and revealing and systems of integrative complexity. *Journal of Personality and Social Psychology*, 1966, **3**, 655–664.

TULKIN, S. R. Author's reply: Environmental influences on intellectual achievement. *Representative Research in Social Psychology*, 1970, **1**, 29–32.

VONDRACEK, F. W. Behavioral measurement of self-disclosure. *Psychological Reports*, 1969, **25**, 914. (a)

VONDRACEK, F. W. The study of self-disclosure in experimental interviews. *Journal of Psychology*, 1969, **72**, 55–59. (b)

VONDRACEK, F. W., & MARSHALL, M. J. Self-disclosure and interpersonal trust: An exploratory study. *Psychological Reports*, 1971, **28**, 235–240.

VONDRACEK, S. I., & VONDRACEK, F. W. The manipulation and measurement of self-disclosure in preadolescents. *Merrill-Palmer Quarterly*, 1971, **17**, 51–58.

WEICK, K. E. The "ess" in stress: Some conceptual and methodological problems. In J. McGrath (Ed.), *Social and psychological factors in stress*. New York: Holt, Rinehart & Winston, 1970.

WEIGEL, R. G., WIEGEL, V. M., & CHADWICK, P. C. Reported and projected self-disclosure. *Psychological Reports*, 1969, **24**, 283–287.

WEST, L., & ZINGLE, H. W. A self-disclosure inventory for adolescents. *Psychological Reports*, 1969, **24**, 439–445.

WORTHY, M., GARY, A. L., & KAHN, G. M. Self-disclosure as an exchange process. *Journal of Personality and Social Psychology*, 1969, **13**, 59–63.

(Received November 4, 1971)

Psychological Bulletin
1973, Vol. 79, No. 2, 92–106

DEVIANT DRUG USE IN ADOLESCENCE:

A REVIEW OF PSYCHOSOCIAL CORRELATES[1]

G. NICHOLAS BRAUCHT[2]

University of Denver

DANIEL BRAKARSH AND DIANE FOLLINGSTAD

University of Colorado

K. L. BERRY

University of Denver

Research concerning the sociocultural and personality correlates of deviant adolescent use of alcohol, psychedelic drugs, and narcotics was reviewed. A number of consistently replicated findings regarding both personality and sociocultural correlates of adolescent use were found. At the same time, lack of comprehensive, systematic and theory-bound research, gross lack of methodological rigor, and a lack of coordination of effort and integration of findings among the various areas of drug use research were noted. General requirements for the design of future research were also discussed.

Deviant adolescent drug use has come to occupy center stage among the long list of contemporary problems concerning America's youth. Given the current drive for intervention into this widespread and seemingly increasing deviant use of drugs (including alcohol) by today's youth, the need for psychosocial explanation of adolescent drug use is of pressing practical concern. Comprehensive, cogent identification of the sociocultural and personality determinants of various forms of youthful drug use is fundamental to the development and administration of primary prevention and treatment efforts in this area. Accordingly, it is the authors' intention to review critically the literature concerning the sociocultural and personality correlates of deviant adolescent drug use in an attempt to assess the level of psychosocial description and explanation which is presently available.

This paper considers deviant adolescent drug use to be defined as any nonmedical use of consciousness-altering drugs by adolescents, and also what has been described (but not precisely defined) as adolescent problem drinking of beverage alcohol. The collection of the review material was facilitated by the gross categorization of the material into subareas of research corresponding to types of drugs, that is, alcohol research, narcotics research, and psychedelic drug research, and this paper is organized by these divisions. For each type of drug use, the material is further divided into research concerning the sociocultural correlates and then the personality correlates of use.

SOCIOCULTURAL CORRELATES OF ADOLESCENT ALCOHOL USE

By the time the adolescent graduates from high school, it is quite likely that he has established a regular (light to moderate) drinking pattern. The proportion of adolescent drinkers varies from community to community—anywhere from 30%–40% to 60%–80% of the adolescent population. This variation reflects regional and ethnic subcultural differences (Maddox, 1970). While the vast majority of adolescents are either abstainers or light to moderate drinkers, somewhere between 2%–6% of the adolescent population are problem drinkers—drinkers for whom alcohol has come to occupy a more than simply incidental role in their lives.

---

[1] Completion of this paper was partially supported by Grant 1 R01 MH 20850 from the National Institute of Mental Health, United States Public Health Service.

[2] Requests for reprints should be sent to G. Nicholas Braucht, Department of Psychology, University of Denver, Denver, Colorado 80210.

The consensus of a large number of empirical studies reviewed by Maddox (1970) was that the majority of adolescent drinkers report that their first drinking experience was in their own home with parents and/or relatives present, that they see drinking as an adult behavior, and that alcoholic beverages are typically readily available around the house. Jessor, Graves, Hanson, & Jessor (1968) found that unsanctioned introductions to drinking and heavy consumption patterns among high school age children were related to the heavy drinking behavior of adult Ute males. In his clinical observations of adolescent problem drinkers, McKay (1961) noted that almost all the fathers and some of the mothers in his sample were alcoholics. Gusfield (1970) found that within a college student sample, high users of alcoholic beverages were more likely than low users to have parents who were also high users of alcohol. Summarizing the research in this area, Maddox (1970) concluded that knowledge of parents' drinking patterns is the single most accurate tool for predicting adolescent drinking behavior. Unfortunately, however, there is a lack of systematic investigation into individual difference variables which could explain the substantial number of false positives in the research reviewed above (e.g., personal feelings of shame over heavy parental drinking could explain why a given adolescent is an abstainer, even though his parents are alcoholics.

More recently, Jessor, Collins, and Jessor (1972) found that the opportunity to learn and to be reinforced for drinking, especially by peers, plays a prime role in the prediction of abstainer to drinker status shifts among junior–senior high school students. Studies with college student peer groups, specifically fraternities, have supported the notion that peer influences on adolescent attitudes toward drinking and drinking behavior are substantial. Gusfield (1970) and Rogers (1970) both found that fraternity members are more likely to be drinkers than nonfraternity members. Gusfield went one step further by showing that fraternity members with high-user parents (who would, therefore, be expected to be high users) drank no more frequently or excessively than fraternity members with low-user parents. Although the drinking occurring in the context of another kind of peer group, adolescent gangs, may at times be an expression of rebellion and hostility, Maddox (1970) has found that this is not typically the case. In the absence of data on the process of influence, findings such as those above are left equivocal. Uncontrolled selective factors (both self and group determined) that affect fraternity or gang membership may play roles which are as significant or more significant than group membership itself.

The evidence on the relation between religious affiliation and drinking behavior is fairly conclusive. Maddox (1970) points out that, due to the norms concerning drinking in the respective religions, the average drinker will probably be a Jew or a Catholic rather than a Protestant. Snyder (1958) found, however, using a sample of college alcohol users, that the intoxication rates for ascetic Protestant and Mormon students is much higher than that of Jewish students. Similarly, Skolnick (1958) found that the rate of social complications due to drinking for Jewish students who drink is much lower than the rate of social complications for student drinkers who are members of ascetic Protestant groups. Finally, Gusfield (1970) found that Protestant college students were more often high users of alcohol than were Jewish students. Thus, while fewer Protestant and Mormon than Jewish students drink, those Protestant and Mormon students who drink are more likely to be problem drinkers than the Jewish students who drink.

Mizruchi and Perrucci (1970) explain these findings by using the concepts of prescriptive and proscriptive norms. The norms in the Jewish religion concerning drinking behavior are prescriptive: there is an elaborate system of explicit directives as to what, when, where, with whom, how much, and why one is expected to consume alcoholic beverages (including, we may add, occasions when drinking to moderate excess is socially tolerable, or even approved). The norms of the Protestant (especially ascetic Protestant) religion are proscriptive in nature: Drinking is forbidden. Thus, when a Jew decides to use alcoholic beverages, norms advocating moderation exert a controlling influence on his drinking behavior. However, when persons from a Mormon or Protestant background use alcoholic beverages, there is an absence of norms to guide their use, and the drinking is more likely to be uncontrolled. The issue is not unrelated to the possible positive and negative roles of the parental model since parents are, after all, the principal early-life transmitters of societal norms.

Two other important points concerning religion's influence upon drinking behavior are made by Maddox (1970): (*a*) Close association with any organized religion will depress the rate of problem-related drinking, as there is no (major) religion which advocates the intemperate use of alcohol; (*b*) ". . . there is no indication that religious organizations as socializing agencies can or do intervene effectively to increase the prevalence of abstinence among persons whose interpersonal relationships otherwise support drinking [p. 116]."

In summary, the research literature concerning the sociocultural correlates of adolescent problem drinking focuses on three interrelated variables: parental models, peer models and sanctions, and religious affiliation. The most common variety of adolescent problem drinker would seem to come from a home where the parents are high users of alcohol and would seem to belong to peer groups in which high alcohol use is also encouraged. Furthermore, such individuals likely have rather loose religious affiliations with one of the abstinence-oriented religions. Unfortunately, however, there has been little reported research into the relationship of adolescent problem drinking to such basic sociocultural variables as a race, socioeconomic status, and living conditions.

PERSONALITY CORRELATES OF ALCOHOL USE

Within the past 10 years there have been many studies which have addressed themselves to the question of whether or not there exists a pattern of personality traits common to most adolescent problem drinkers. The most comprehensive study was conducted by Williams (1970), who used the Adjective Check List as his primary assessment instrument. His results characterize the college-age, male problem drinker as a comparatively independent, aggressive, anxious, impulsive and depressed individual, who deemphasizes both primary and secondary relationships, is relatively unconcerned about others and does not think too highly of himself. In an earlier study, using the same measuring instruments an an all male college population, Williams (1968) found a significant positive correlation between the problem drinking and anxiety.

Williams' (1970) findings, that problem drinkers are relatively impulsive, aggressive, and depressed, have also been noted by MacKay. MacKay (1961) reported many incidents of "impulsive angry outbursts" on the part of juvenile problem drinkers and their parents, and cited evidence of depression in these adolescents ranging from "feelings of irritability, worthlessness, fear of the future . . . poor appetite, strange eating habits, and sleeplessness to overt suicide attempts [p. 26]."

The findings of Jones (1968, 1971) are very similar to those just reviewed. Jones' work is unique in two respects: It is longitudinal in nature, and she uses both male and female subjects. In her longitudinal study of problem drinkers, utilizing preproblem data, Jones found both her male and female preproblem subjects to be more unstable, unpredictable and impulsive than other premoderate drinkers or abstainers. At the same time, she found that females who are destined to be problem drinkers are more unlike female prenormal drinkers than are male preproblem drinkers when compared to male prenormal drinkers.

Jessor, Carman, and Grossman (1968) found that individuals in a college environment who have low expectations of attaining academic and social recognition will tend to drink more often, get drunk more often, and get into trouble over drinking more often than will those individuals who have higher expectations in these two areas.

Findings in the area of sexual identity are quite confusing. Parker (1969) found that a sex-role conflict exists for male college students with a strong alcohol disposition. They want to be perceived as very masculine, but at the same time they themselves do not identify with masculine traits. A study by Zucker (1968) concluded that adolescent problem drinkers are not experiencing a sex-role conflict, but are only attempting to present to the world a sex-role facade—high school problem drinkers only want to *seem* more masculine than do the moderate to light drinkers. The disparity in Parker and Zucker's results may be due to the use of different samples (college, high school), or to ambiguities in the concept of sexual identity and the derivative inconsistencies in the criteria.

While the above research suggests a number of preproblem use correlates of problem alcohol use, why, given these person states, does the youth turn to alcohol use? Only Jessor, Carman, and Grossman (1968) and Williams (1966, 1968) consider this last question to any significant extent. Jessor, Carman, and Grossman hypothesize that individuals in a college setting will become highly involved with alcohol when their expectations of attaining two socially and presumably personally valued goals are low. In this context, heavy alcohol consumption is seen as an alternative way of attaining these previously unattainable goals, a way of attaining substitute goals, or a way of coping with actual or anticipated failure in regard to goal attainment. Jessor, Carman, and Grossman showed that a relationship does exist between low expectations of need satisfaction (a need here is viewed as an internalized goal) and high alcohol involvement.

It is a widely held belief that many adolescent problem drinkers drink and get drunk often in order to escape from their high levels of anxiety and depression. Williams (1966) tested this idea and found that alcohol does afford problem drinkers with some relief when they are moderately intoxicated, but not when they are severely intoxicated. While the results of his study can explain why an anxious and depressed person might drink moderately and frequently, they do not help to explain why an anxious and depressed individual would become a problem drinker (associated with both excessive and frequent drinking) since relief from anxiety and depression is apparently not gained from excessive alcohol consumption.

Using the personality variables which differentiated problem drinkers from nonproblem drinkers in his 1970 study, Williams (1968) studied personality changes occurring as a person consumes more and more alcohol. Using a college male sample, Williams found that the personality descriptions of nonproblem drinkers under the influence of alcohol closely matched the personality of problems drinkers when sober. Williams concludes that a problem drinker feels more at home in a drinking situation than a nonproblem drinker because he is around people who are now aggressive, self-centered, impulsive and the like, just as he is. This would explain why an adolescent male with this given set of personality traits might tend to gravitate toward situations where alcohol is being served and tend to drink frequently, but it does not explain why he will drink excessively.

McClelland's (1971) study attempted to answer the question, "What are the psychological effects of alcohol?" McClelland found that as male students at fraternity parties drank more, there was a corresponding increase in power thoughts (as indicated on a Thematic Apperception Test type measure—thoughts of having impact on others, thoughts of aggressions, of sexual conquest and of being strong, big, and influential. After moderate drinking these power thoughts centered on the altruistic use of power on behalf of others, while after heavy drinking the power thoughts concerned personal dominance over others. McClelland concluded that the problem drinker is an individual who is intensely concerned with his own personal power and who has chosen drinking as a way to accentuate his feelings of power.

In summary, there appears to be substantial agreement that adolescent problem drinkers are individuals lacking in personal controls, as evidenced by relatively high aggressiveness and impulsiveness. At the same time, there exists some evidence which indicates that adolescent problem drinkers have some basic neurotic tendencies which form a coherent cluster of traits: relatively low self-esteem, high anxiety, depression, and a general lack of success in the attainment of life goals.

There are some basic areas of disagreement concerning the existence of two basic conflicts within the personality of the adolescent problem drinker: independence and sex-role conflicts. A likely contributor to this problem is the use of different measures of criteria (e.g., sex-role conflict; Parker, 1969; Zucker, 1968). Tied to this issue is the questionable rationale for criterion choices made by many of the researchers. Although the independence–dependence trait, for example, may be correlated with adolescent problem drinking, there is no compelling theoretical rationale for the choice of this trait as a prime independent variable. The Jessor, Carman, and Grossman (1968) study is the exception to this rule as it is made conceptually clear, prior to any empirical research, why alcohol use as a coping behavior is related to low expectation of goal attainment. Furthermore, most research on alcohol consumption among adolescents has been limited to males of college age. It is not at all certain that these results can be generalized to college females, or to males or females who are under college age.

SOCIOCULTURAL CORRELATES OF ADOLESCENT NARCOTIC USE

Much research has focused on the familial environment of the adolescent addict. Family fragmentation, in the form of either the actual physical loss (Vaillant, 1966b) or the separation of parents (Rosenberg, 1968) has often been noted in the family backgrounds of adolescent addicts. These findings are further supported by those of both Rosenberg (1969) who pointed out that less than half of his subject group of youthful addicts had reached the age of 15 with both their parents living continuously at home, and by Lewis and Osberg (1958) who noted that the family of the addict generally lacks cohesiveness.

Although many researchers agree that the parents of addicts are extreme personality types (Modlin & Montes, 1964), there is a marked lack of agreement as to the nature of the extreme personality type. Three major parental personality patterns have emerged in the literature. The first is the overprotecting parent, found in the research of Ausubel (1961), Bender (1963), and Laskowitz (1961). The second pattern is that of the underdominating parent (Ausubel, 1961; Bender, 1963). This personality pattern is supported by Rosenberg's study (1968) which discovered a relationship between inadequate parental control and adolescent addiction and the research of Robins (1966), which found that subjects who were diagnosed as psychopaths having problems with alcohol and drugs were early in the sibling birth order and frequently came from families whose parents were uninterested in them and exercised little discipline. While Rosenberg (1969) has isolated the trait of underdominance as typical of the mother of the adolescent addict, Lewis and Osberg (1958) hypothesize that the father, if not absent, is the weak individual in the family.

The third parental personality pattern thought to be related to adolescent addiction is overdomination (Ausubel, 1961; Bender, 1963). Zimmering (1952) and Lewis and Osberg (1958) feel that the pattern consists specifically of an overdominating mother, while Mason (1958) posits that these overpowering mothers simulate aggression and sexual drives in their children, then deny and punish these drives. Rosenberg (1969) found that the father of the adolescent addict represented the overdominating influence in the family, frequently the symbol of punishment and rejection.

Another theme in the research on sociocultural correlates of adolescent addiction is parental rejection during the formative years. While Rosenberg (1969) points to frequent paternal rejection, Hirsh's (1961) theory suggests that sons who are infantilized by rejecting narcissistic mothers are prone to addiction. Gold (1957) also suggests the salience of a rejecting environment within families, producing insecurity feelings. Rosenberg (1968, 1969) concludes that the early life of the addict is marked by either a general deprivation of parental models or the presence of inadequate models, as indicated by a high degree of parental alcohol abuse or mental or physical illness. Nyswander (1956) disagrees to some extent with the deviant family model by stating that addicts have no more or less problems with their parents than the average emotionally disturbed patient. Although there is a lack of consensus as to specifics in the studies noted above, this body of research (primarily retrospective studies) seems to point to the hypothesis that the adolescent addict, in his early years, suffers from mentally disturbed parents, from rejecting parents, or from the absence of one or both parents.

The ethnic composition of adolescent narcotic addicts has been studied by Chein (1959), Bender (1963), and Vaillant (1966a), who agree that most addicts are members of a minority community. There seem to exist two implications of this finding. First, being a member of an ethnic minority leads to the possibility of racial discrimination; both Murray (1967) and Zimmering (1952) found racial discrimination to be a relevant factor in narcotics addiction. Second, Vaillant's (1966a) sample of adolescent addicts were not only from a minority community but were first generation, native born Americans. This finding points to the possibility that one of the sociocultural factors relevant to narcotics addition is a cultural clash between parent and child.

Chein (1959), Bender (1963), and Vaillant (1966b) found that narcotic addicts are generally found in urban and metropolitan areas. These addicts usually grow up in poor residential areas (Murray, 1967) and come from families in low-income groups (Bender, 1963; Chein, 1959; Vaillant, 1966b).

The hypothesis that there is some parent–child conflict regarding status achievement is supported by the studies of Rosenberg (1968, 1969), Vaillant (1966b), and Laskowitz (1961) who found an inadequate vocational adjustment history in the backgrounds of their addict sample. The studies of Ausubel (1961), Scher (1966), Gamso and Mason (1958), Cameron (1963), and Modlin and Montes (1964) also indicate that it is in these urban, poor residential areas that there is a relatively high availability of narcotics, and it is this availability which also helps to determine who becomes a drug addict.

A number of studies (Cameron, 1963; Gamso & Mason, 1958; Modlin & Montes, 1964; Scher, 1966; Vaillant, 1966b; have shown that adolescent narcotics addiction is usually connected with peer-group enticement to deviance. Furthermore, Little and Pearson (1966) attribute perpetuation of drug addiction to the interdependence of an addict and others whom the addict persuades to experiment with narcotics.

In summary, research indicates that the adolescent narcotic user is frequently a members of an ethnic minority, often from an impoverished urban environment, often from a broken home. There is further consensus regarding the importance of peer group enticement and availability of narcotics. While there seems to be agreement on the general point that adolescent users are usually subject to deficient parental models, there is marked disagreement as to the nature of this deficiency; one or both of the parents are characterized as being either overprotecting, overdominating, underdominating, or rejecting. This lack of consensus can be attributed to the heavy reliance on retrospective studies of the narcotic users' parents.

PERSONALITY FACTORS RELATED TO NARCOTIC USE

Intelligence has been a variable considered by several researchers. Zimmering (1952) found that most addicts to be of dull normal intelligence, but Rosenberg (1968), Vaillant (1966b), and Knight and Prout (1951) disagree. They found their subjects to range from average to superior intelligence. Immaturity is considered a personality characteristic of youthful addicts by Rettig and Pasamanick (1964), Ausubel (1961), and Bender (1963). Both Bender (1963) and Zimmering (1952), even though their observations were conducted a decade apart, agree that addicts show low frustration tolerance, that they tend to repress their troubles, and that they tend to withdraw into fantasies. Knight and Prout (1951) described their clinic population of addicts as being intoverted, shy, insecure (also Gilbert & Lombardi, 1967), having a flat affect, shallow interests and immature goals, and lacking healthy resources and adequate structure. Gilbert and Lombardi (1967), using Minnesota Multiphasic Personality Inventory data, concluded that addicts generally are depressed (also Leeds, 1965), tense, feel inadequate (also Laskowitz, 1961), are irresponsible (also Laskowitz, 1961), are impatient and irritable, lack persistence, are hypersensitive and apprehensive, and have poor morale. Both Gilbert and Lombardi (1967) and Laskowitz (1961) noted the egocentrism of addicts in their disregard for social mores and in their view of themselves as members of an elite group. Restlessness and quick dissatisfaction, private logic following, limited problem solving, and overvaluation of the masculine role are described by Laskowitz (1961). Savitt (1963) describes the addict as a person unable to experience love and gratification through the usual channels. An interesting study by Cooper (1959) suggests that the personality descriptions of drug addicts in literature are not at all like the self-perceptions of these individuals.

Most researchers agree that most narcotic addicts have some mental problems or a weak or disturbed personality (Ausubel, 1961; Chein, Gerard, Lee, & Rosenfeld, 1964; Gerard & Kornetsky, 1954; Holmberg & Jansson, 1968; Laskowitz, 1961; Savitt, 1963; Vogel, Isbell, & Chapman, 1948; Wakefield, 1963). The consensus ends here. Many authors suggest their own theories and findings as to the mental status of addicts: Cameron (1963) describes addicts as inadequate personalities; Sorio (1967) discusses a pathological desire for drugs; Rosenberg (1968) characterizes his subjects as having high levels of anxiety and neuroticism; and Laskowitz (1961) also stresses the neurotic tendencies of addicts, considering them to have low ego strength or poorly developed superegos. Vaillant (1966b) and Bender (1963) call the syndrome a personality disorder, Gilbert and Lombardi (1967) and Cameron (1963) call addiction a basic character disorder, while Murray (1967) and Laskowitz (1961) use the term social deviant, and Hill (1957) employs the term conduct disorder. Psychopathic traits are considered to be present in many addicts by Gilbert and Lombardi (1967), Hill (1957), Monroe, Miller, and Lyle (1964) and Laskowitz (1961). Hill, Haertzen, and Glaser (1960) used Minnesota Multiphasic Personality Inventory data to conclude "Personality characteristics of narcotic addicts are either associated with psychopathy or are predominately psychopathic in nature although they may include many of the classical psychoneurotic and psychotic features [p. 97]." Rosenberg (1969) reported less disturbance on psychological tests by addicts than alcoholics, although Gilbert and Lombardi (1967) found addicts to have deeper and more widespread pathology than nonaddicts. Holmberg and Jansson (1968) consider schizoid curiosity to be a factor in narcotic use, Hill (1957) suggests that some addicts are schizoid psychopaths, and Bender (1963) reports evidences of schizoid blocking of the perceptual field in addict samples. Clark (1962) reports that psychiatric illnesses precede or accompany 40% of drug addiction cases and that 20% of the cases appear to be sociopathic. Nyswander (1956) has found that the reported incidence of mental illness among addicts is no different from the rate in the general population, and she suggests that this is the result of addicts' failure to seek or avail themselves of help, and because their psychopathology is often subclinical.

In summary, addicts have been found to be immature, insecure, irresponsible, and egocentric. However, the way in which all these traits fit together within any given narcotic addict is unclear. This can be seen as the result of a lack of a sound theoretical structure as the point of origin for these studies. Furthermore, although there is some agreement that the narcotic addict suffers from a personality disorder, there is hardly any agreement as to the specific dynamics of this disorder. Indeed, one should be aware that the population of adolescent drug addicts is very loosely defined and includes more than one personality type. Psychological predispositions to recruitment into this population may be different for different sociocultural strata, different patterns of drug use (e.g., chronic or episodic) and differential involvement of various psychological functions of drug use. On this last point, see Chein et al. (1964) and Chein (1969).

Almost all of the personality research and much of the sociocultural research focuses on the psychopathological nature of adolescent narcotic use. Although such an approach is reasonable if the goal is the treatment of the young addict, it is not readily apparent that the psychopathology orientation is appropriate to the primary prevention of narcotic use. In addition, the use of uncontrolled convict and clinic populations in this body of research casts considerable doubt on the validity, reliability, and representativeness of the research findings. Furthermore, the personality results reported are primarily the product of small sample observational research, or in many cases, conjecture.

SOCIOCULTURAL CORRELATES OF ADOLESCENT PSYCHEDELIC DRUG USE

Most of the research into psychedelic drugs has been devoted to studies of marijuana, LSD, and methamphetamine—three distinct drugs that will be subsumed here under the psychedelic label.

With respect to the gross group definers of sex and age, Pearlman (1968) found no difference between psychedelic users and nonusers in his college population. Harris (1971), however, found males more likely to use marijuana than females. Hager, Verner, and Stewart (1971) found males reporting more experience with all types of drugs, although both sexes showed similar patterns of consumption. Mitchell, Kirkby, and Mitchell (1971), in a study done with college freshmen, found different use patterns between males and females. Hager et al. (1971) found a pattern of drug use among adolescents in which there was a definite increase in use with age. As usual, contradictory evidence is available—Mizner, Barter, and Werme (1970) found that younger college-age students fell into moderate or heavy use categories while older students, even though they presumably have had more opportunity for being acquainted with drugs, have a more experimental or casual use pattern. Their sample may have been biased, however, as heavy users may have dropped out of college and thus may have been underrepresented.

A number of studies have indicated that the social class of psychedelic users is generally middle to upper middle class in significant contrast to narcotic addicts (Carey & Mandel, 1968; Gillie, 1969; Hager et al., 1971; Harris, 1971; Kuehn, 1970; Smart & Fejer, 1969; Steffenhagen, McAree, & Zhevtlin, 1969; Suchman, 1968; Welpton, 1968). This social class distinction is consistent for marijuana users as well as users of the more powerful psychedelic drugs, that is, LSD and methamphetamines. Users of LSD commonly have fathers who are either professional or managerial and have parents with high levels of education (Smart & Fejer, 1969).

Although McGlothlin and Cohen (1965) feel that atheists, agnostics, and believers in pantheistic religions would more likely be users than Protestants, Catholics, or Jews, Steffenhagen et al. (1969) found 70% of their volunteer college sample to be Jewish. Finally, Dearden and Jekel (1971) found church attendance in a high school sample to be inversely associated with drug use.

Steffenhagen et al. (1969) reported the families of marijuana users as not being disorganized and disrupted. In contrast, Harris (1971) found marijuana users often had early childhood rearing with other than both parents, while Gillie (1969) reports that the percentage of parental loss is higher for psychedelic users than for nonusers. Through observation, Kuehn (1970) found families of psychedelic drug users to be characterized by excessive permissiveness, material gratification with inadequate or inappropriate limit-setting, parents with histories of severe emotional instability, family conflict suppressed behind a facade, communication binds, parental love more talked about than provided, parental role reversals, and a lack of tolerance for anger. Welpton's (1968) study of LSD users suggests that the parents of users have failed in their efforts to have an intimate relationship with one another. Keniston (1965) described mothers of psychedelic users as talented women who "gave it all up for marriage" and described the fathers as either having failed in life or "sold out" for success. An important problem with most of these kinds of studies lies in their postdictive and retrospective qualities.

A close knit group which is highly involved in psychedelic drug use has been noted as being a critical aspect of the peer subculture by a number of researchers (Freedman, 1968; Klein & Phillips, 1968; Ludwig &

Leving, 1965; Mamlet, 1967; Scher, 1966; Welpton, 1968). Norton (1968) spoke of the narcissistic aspect of the group process as being more important than marijuana itself. Carey and Mandel (1968) suggested the clannishness and strong bonds of trust and cameraderie in such groups were due to the realization of risks and fears of getting caught. Malcolm (1970) likens the drug subculture to a religious or political cult which attracts the immature, the mentally disturbed, the socially disadvantaged, and the sociopathic, this subculture serving to crystallize the user's socially maladjusted view of reality. The final results become withdrawal from competition, passivity, a rejection of logic and rationalism, immediate gratification of personal urges, and a shared belief in the utter rightness of violence or of ritual execution. Suchman (1968) sees a close association between use of drugs (mainly marijuana) and adherence of a "hang loose" ethic which is indicative of opposition to traditional, established order. Keniston (1965) views drug use as a by-product of the alienation and noncommitment which make up a large part of the drug subculture. Harris (1971) hypothesizes that drug usage is a form of group rebellion against or alienation from the established society. Harris' conclusion—that students who band together in groups characterized by anomic variables of isolation, powerlessness, and normlessness are more likely to use marijuana than those who do not—does not appear to be a firmly established cause–effect relationship, however. McAree, Steffenhagen, and Zheutlin (1969) describe marijuana use as a badge of youthful group protest.

In summary, the research in the area of sociocultural correlates of adolescent psychedelic use has centered on six major variables: sex, age, socioeconomic status, familial environment, peer group influence, and religious values. While there is considerable agreement concerning the findings that adolescent psychedelic users come mainly from the middle and upper classes, this is where the consensus terminates. The research regarding the role played by the five other sociocultural variables is plagued by either ambiguous or contradictory findings. As a result, no firm conclusion can be drawn as to the sociocultural correlates of adolescent psychedelic use, with the exception of the socioeconomic status findings.

Personality Correlates of Psychedelic Use

Edwards, Bloom, and Cohen (1969) found considerable hostility among heavy drug users, perhaps reflecting a predisposing state, as amount of hostility was unrelated to the degree of drug dependence. Cockett and Marks (1969) found higher total hostility scores among amphetamine users in a juvenile delinquent sample. Other personality characteristics reported by these authors were that users are more intropunitive (self-criticism and guilt), have higher anxiety, exhibit lower extraversion, are more shy and retiring, less self-confident, and more radical than traditional.

Kuehn (1970) observed 41 college students and suggested the following eight traits for student drug users: (*a*) tendency to live excessively in the present; (*b*) an excessive passive and reactive position in interpersonal relationships; (*c*) serious cognitive difficulties (schizophrenoid disturbances in thinking and verbalization); (*d*) inexplicable depression; (*e*) study difficulties not attributable to difficulties in the environment; (*f*) unrewarding sexual behavior; (*g*) use of repression and rationalization; (h) intellectualization and isolation as secondary defenses.

In a study with middle-class high school students, Green, Blake, Carboy, and Zenhausenn (1971) found that marijuana users as a whole tend to be more vulnerable to frustration, more headstrong, reckless, group dependent, and less self-controlled. The same study also found two distinct types of marijuana users, distinguished on the basis of academic performances. The high-performance user was found to be intelligent, alert, confident, and sensitive, while the low-performance user was depicted as being shy, cynical, unstable, and less able to handle abstract thinking.

Hogan, Mankin, Conway, and Fox (1970) collected California Personality Inventory data from college undergraduate marijuana users and nonusers. The users were found to be socially poised, open to experience, concerned with feelings of others, self-confident, skilled in interpersonal relations, socially perceptive, to possess a wide range of interests, and to have the achievement motivation necessary for success in graduate school. Marijuana users were also found to be impulsive, pleasure-seeking, rebellious, tending toward narcissism and self-aggrandizement, overly concerned with personal pleasure and external diversion, hostile to roles and conventions, somewhat irresponsible, and rather nonconforming. Nonusers appeared to be responsible, rule abiding, pleasant, considerate and dutiful, but they also tended to be conventional, lacking in spontaneity, flexibility, and verve, perhaps too deferential to external authority, narrow in their interests and overcontrolled.

In the Hogan et al. (1970) study, users were found to major mostly in the humanities and social sciences, and nonusers exhibited slightly poorer grades than users. Results of other studies have characterized psychedelic users as being underachievers (Smart & Fejer, 1969), as being indistinguishable from nonusers (Pearlman, 1968), and as being slightly above average academically (Steffenhagen et al., 1969). In a population of youthful offenders using amphetamines, Cockett and Marks (1969) found no association between amphetamine-taking and intelligence level. Cohen and Klein (1970) found that users in a hospital population were of higher intelligence than nondrug users. One uncontrolled variable that could account for this is that persons of higher intelligence may be more apt to seek out unusual (i.e., psychedelic) experiences or that they may be more apt to seek hospitalization.

In an attempt to determine the psychopathological correlates of psychedelic use, McAree, Steffenhagen, and Zheutlin (1969) found that marijuana-only users were not significantly different from controls on Minnesota Multiphasic Personality Inventory data (in some contrast to Hogan et al., 1970). Multiple-drug users, however, showed high ranking on the schizophrenia scale (representing withdrawal, poor interpersonal relationships, aloofness and the inability to express emotions). They also scored higher on the Masculinity–Femininity scale (noted as possibly due to increased sensitivity and aesthetic appreciation), the Psychopathic Deviate scale and the Hypomania scale. Brill, Compton, and Grayson (1971) also found that college undergraduate marijuana users (some of whom used other drugs) scored significantly higher than nonusers on the Minnesota Multiphasic Personality Inventory Psychopathic Deviant scale. On the other hand, Davis and Brehm (1971) in a study of young drug users in the North Carolina State Prison system, found no significant difference between users and nonusers. Users in this study consisted of users of both psychedelics and hard drugs.

McAree, Steffenhagen, and Zheutlin (1969) hypothesize that the potentially more disturbed individual is attracted toward the more dangerous forms of drugs. Abnormal Minnesota Multiphasic Personality Inventories were found in the LSD users in the study by Smart and Fejer (1969), and diagnoses of personality disorder or borderline psychosis were suggested. Frosch, Robbins, and Stern (1965) support these findings with a sample of 12 LSD patients diagnosed as psychotic or borderline psychotic. Blumenfield and Glickman (1967) also surveyed a hospital population of youthful LSD users in which over 70% had had previous psychiatric treatment and 80% were diagnosed as psychotic or borderline. They concluded, however, that the patients' difficulties antedated the use of LSD (see also Flynn, 1970). Cohen and Klein (1970) in reviewing a private hospital population, found heavy drug users more likely to be character disorders than psychotic. They suggest that the divergence of their findings from other studies may be due to the fact that users admitted to public hospitals are more likely to be psychotic (or *diagnosed* psychotic). The clinic population studied by Hekimian and Gershon (1968) revealed that 50% of the psychedelic-using patients had been considered schizophrenic or treated for schizophrenia before drug ingestion. In marked contrast to the diagnosis of narcotic users, Hensala et al.

(1967) found that discharge diagnoses for multidrug users indicated more schizophrenic disorders and fewer personality disorders among the drug group than controls. The general consensus concerning mental illness and use appears to be that persons with predisposing psychopathology often take drugs which aggravate that state. Use of drugs themselves is not seen as producing abnormality.

The foregoing suggests a multitude of preproblem use correlates of adolescent psychedelic use. At the same time, there is very little consensual validation and there are many contradictory findings. Even if all of these personality correlates, which are too numerous to list, were found to be valid, two important questions would still be left unanswered. First, how do all these correlates fit together within a given individual? Second, why, given these person states, does an individual turn to psychedelic use?

Sadava (1970) reviews nine themes which may be functions of psychedelic drug use by college students: (*a*) alienation and search for meaning (Blum et al., 1966; Dearden & Jekel, 1971; Freedman, 1968; Keniston, 1965; Kuehn, 1970; Leary, Metzmer, & Alpert, 1964; Pearce, 1971; Welpton, 1968); (*b*) disillusionment and rebellion (Brooks, 1971; Dearden & Jekel, 1971; Janowitz, 1967; Mizner et. al., 1970; Suchman, 1968); (*c*) need for stimulation (Flynn, 1970; Keniston, 1965; Liebert, 1967); (*d*) self-definition (Barron, 1967; Blum, 1966; Brooks, 1971; Davis & Munoz, 1968; Freedman, 1968; Keeler, 1968; Kleber, 1965; Kuehn, 1970); (*e*) interpersonal relationships (Blum, 1966; Bowers, Chipman, Schwartz, & Dann, 1967; Dearden & Jekel, 1971; Freedman, 1968; Keeler, 1968; Kleber, 1965; Mizner et al., 1970; Pearce, 1971); (*f*) escape from sexuality (Blum, 1966; Flynn, 1970); (*g*) hedonism (Blum, 1966; Davis & Munoz, 1968; Flynn, 1970; Kuehn, 1970); (*h*) relief of anxiety and tension (Blum, 1966; Brooks, 1971; Flynn, 1970; Janowitz, 1967; Keeler, 1968; Mitchell et al., 1970; Pearce, 1971); and (*i*) curiosity and novelty (Blum, 1966; Dearden & Jekel, 1971; Flynn, 1970; Janowitz, 1967; Keeler, 1968; Kleber, 1965; Liebert, 1967; Mizner et al., 1970)—a mind-boggling array of potential determinants of psychedelic use.

As with research on narcotic use, there are definite problems in obtaining data which are valid, representative, and reliable. Due to the illegality of psychedelic drugs, voluntary populations are difficult to obtain and accuracy of information is often questionable. The heavy use of college populations biases samples through exclusion of ghetto groups and other groups who use the same drugs but may possess different personality correlates or motivations for use. Other common sources of subjects also produce questionable results—the uncontrolled use of clinic and incarcerated delinquent populations. Almost none of the research is based on high-school-aged or younger populations—populations in which primary prevention is more feasible. Also, the near total reliance on male samples makes untenable the generalization of the findings to females. Finally, as with previously summarized research, the use of small-sample observational studies and empirically unsupported theoretical offerings is extensive.

CONCLUSION

Within the summaries at the end of each section above, a number of firm conclusions regarding the psychosocial correlates of adolescent drug use have been detailed. However, the body of reviewed research suffers from a number of methodological and theoretical shortcomings.

The widespread use of unrepresentative, uncontrolled subject samples and small-sample observational data make it difficult to delineate the salient personality correlates of adolescent drug use. Much of the sociocultural research is either retrospective or based upon small-sample observational studies. Furthermore, very little longitudinal research has been reported in this area.

The research reviewed also suffers from a number of more substantive problems. First, the correlates under study in most of the reviewed research are not embedded in a theoretical framework which is focused on a psychosocial explanation of adolescent drug use. Thus, many variables which are intuitively appealing have been studied in isolation from other variables, and a large number of isolated statements have resulted. For these statements to be of explanatory value, however, their relationship as regards use and their place in the development of deviant drug use must be made compellingly clear by being embedded in an explanatory theoretical structure.

Another related problem with the research is the general lack of comparability and coordination of the sociocultural and personality lines of inquiry, and the consequent lack of control for personal factors in sociocultural studies, and vice versa. If it is reasonable to assume that it is the interaction of the person and his sociocultural environment, and not either alone, that results in a specific behavior—be it studying for an exam or smoking marijuana—then it is not enough to say, for example, that drug use is the result of low personal expectations for success, or that it is the result of models and sanctions for such use. Rather, it is the combination of these that should offer greater prediction and explanation. What is required for an adequate *psychosocial* understanding of drug use is a theoretical approach to research embracing both sociocultural and personality variables.

A final critical issue posed by the research, reflected in the format of this review, is the lack of coordination among alcohol, narcotic and psychedelic researchers. The same variables may be involved in each kind of use, and therefore, there should be coordinated examination of all kinds of adolescent drug use along theoretically salient social–psychological dimensions. For example, the adolescent narcotic user and alcohol user might be found to differ merely in degree of alienation (all other things being equal).

In conclusion, the formulation of adequate research designs investigating the sociocultural and personality correlates of adolescent drug use requires a comprehensive theoretical system which takes into account sociocultural and personality factors and relates them to the development of deviant drug use. With such a theoretical base and with commensurate methodological sampling and statistical sophistication, empirical investigations could be made which do justice to the complexity of the phenomenon of deviant adolescent drug use.

REFERENCES

Ausubel, D. P. Causes and types of narcotic addiction: A psychosocial view. *Psychiatric Quarterly*, 1961, **35**, 523–531.

Barron, F. Motivational patterns in LSD usage. In R. C. DeBold & R. C. Leaf (Eds.), *LSD, man, and society*. Middletown, Conn.: Wesleyan University Press, 1967.

Bender, L. Drug addiction in adolescence. *Comprehensive Psychiatry*, 1963, **4**, 131–134.

Blum, R. *Utopiates*. New York: Atherton Press, 1966.

Blumenfield, M., & Glickman, L. Ten months experience with LSD users admitted to a county psychiatric receiving hospital. *New York State Journal of Medicine*, 1967, **67**, 1849–1853.

Bowers, M., Chipman, A., Schwartz, A., & Dann, O. T. Dynamics of psychedelic drug abuse. *Archives of General Psychiatry*, 1967, **16**, 560–566.

Brill, N. Q., Compton, E., & Grayson, H. M. Personality factors in marijuana use. *Archives of General Psychiatry*, 1971, **24**, 163–165.

Brooks, H. B. Teaching teachers to teach about drugs. *National Association of Secondary School Principals Bulletin*, 1971, **55**, 127–134.

Cameron, D. C. Addiction: Current issues. *American Journal of Psychiatry*, 1963, **120**, 313–319.

Carey, J. T., & Mandel, J. A San Francisco Bay Area "speed scene." *Journal of Health and Social Behavior*, 1968, **9**, 164–174.

Chein, I. The status of sociological and social psychological knowledge concerning narcotics. In, *Narcotic drug addiction problems*. Washington, D. C.: United States Department of Health, Education, and Welfare, 1959.

Chein, I. Psychological functions of drug use. In H. Sternberg (Ed.), *Scientific basis of drug dependence*. London: Churchill, 1969.

Chein, I., Gerard, D., Lee, R., & Rosenfeld, E. *The road to H*. New York: Basic Books, 1964.

Clark, J. A. The prognosis in drug addiction. *Journal of Mental Science*, 1962, **108**, 411–418.

Cockett, R., & Marks, V. Amphetamine taking among young offenders. *British Journal of Psychiatry*, 1969, **115**, 1203–1204.

Cohen, M., & Klein, D. F. Drug abuse in a young psychiatric population. *American Journal of Orthopsychiatry*, 1970, **40**, 448–455.

Cooper, M. Differences in self-perception among physically dependent drug addicts, alcohol addicts and controls. *Dissertation Abstracts*, 1959, **19**, 2672.

Davis, F., & Munoz, L. Heads and freaks: Patterns and meanings of drug use among hippies. *Journal of Health and Social Behavior*, 1968, **9**, 156–164.

Davis, G. C., & Brehm, M. L. Juvenile prisoners: Motivational factors in drug use. *Proceedings of* *the Annual Convention of the American Psychological Association,* 1971, **6**(Pt. 1), 333–334.

Dearden, M. H., & Jekel, T. F. Pilot program in high school drug education utilizing nondirective techniques and sensitivity training. *Journal of School Health,* 1971, **41**, 118–124.

Edwards, A. E., Bloom, M. H., & Cohen, S. The psychedelics: Love or hostility potion? *Psychological Reports,* 1969, **24**, 843–846.

Flynn, W. R. The pursuit of purity: A defensive use of drug abuse in adolescents. *Adolescence,* 1970, **5**, 141–150.

Freedman, D. S. On the uses and abuses of LSD. *Archives of General Psychiatry,* 1968, **18**, 330–347.

Frosch, W., Robbins, E., & Stern, M. Untoward reactions to LSD resulting in hospitalization. *New England Journal of Medicine,* 1965, **273**, 1235–1239.

Gamso, R. R., & Mason, P. A. A hospital for adolescent drug addicts. *Psychiatric Quarterly Supplement,* 1958, **32**, 99–109.

Gerard, D., & Kornetsky, C. A social and psychiatric study of adolescent opiate addicts. *Psychiatric Quarterly,* 1954, **28**, 113–125.

Gilbert, J. G., & Lombardi, D. N. Personality characteristics of young male narcotic addicts. *Journal of Consulting Psychology,* 1967, **31**, 536–538.

Gillie, O. Drug addiction: Facts and folklore. *Science Journal,* 1969, **5A**, 75–80.

Gold, L. Toward an understanding of adolescent drug addictions. *Federal Probation,* 1957, **22**, 42–48.

Green, M. G., Blake, B. F., Carboy, G. J., & Zenhausenn, R. T. Personality characteristics of the middle class high school drug user. *Proceedings of the Annual Convention of the American Psychological Association,* 1971, **6**(Pt. 2), 559–560.

Gusfield, J. The structural context of college drinking. In G. L. Maddox (Ed.), *The domesticated drug: Drinking among collegians.* New Haven: Conn.: College and University Press, 1970.

Hager, D. L., Verner, A. M., & Stewart, C. S. Patterns of adolescent drug use in middle America. *Journal of Counseling Psychology,* 1971, **18**, 292–297.

Harris, E. M. Measurement of alienation in college students: Marijuana users and nonusers. *Journal of School Health,* 1971, **41**, 130–133.

Hekimian, L. J., & Gershon, S. Characteristics of drug abusers admitted to a psychiatric hospital. *Journal of American Medical Association,* 1968, **205**, 125–130.

Hensala, J., Epstein, L., & Blacker, L. LSD and psychiatric inpatients. *Achives of General Psychiatry,* 1967, **16**, 554–559.

Hill, H. E. How to help the addict. *Contemporary Psychology,* 1957, **2**, 113–114.

Hill, H. E., Haertzen, D. A., & Glaser, R. Personality characteristics of narcotic addicts as indicated by the MMPI. *The Journal of General Psychology,* 1960, **62**, 127–139.

Hirsh, R. Group therapy with parents of adolescent drug addicts. *Psychiatric Quarterly,* 1961, **35**, 702–710.

Hogan, R., Mankin, D., Conway, J., & Fox, S. Personality correlates of undergraduate marijuana use. *Journal of Consulting and Clinical Psychology,* 1970, **35**, 58–63.

Holmberg, M. B., & Jansson, B. Experiences from an outpatient department for drug addiction in Goteberg. *Acta Psychiatrica Scandinavica,* 1968, **44**, 172–189.

Janowitz, J. D. There's no hiding place down there. *American Journal of Orthopsychiatry,* 1967, **37**, 296.

Jessor, R., Carman, R., & Grossman, P. Expectations of need satisfaction and drinking patterns of college students. *Quarterly Journal of Studies on Alcohol,* 1968, **29**, 101–116.

Jessor, R., Collins, M. I., & Jessor, S. L. On becoming a drinker: Social-psychological aspects of an adolescent transition. *Annals of the New York Academy of Sciences,* 1972, **197**, 199–213.

Jessor, R., Graves, T. D., Hanson, R. C., & Jessor, S. L. *Society, personality, and deviant behavior: A study of a tri-ethnic community.* New York: Holt, Rinehart & Winston, 1968.

Jones, M. C. Personality correlates and antecedents of drinking patterns in adult males. *Journal of Consulting and Clinical Psychology,* 1968, **32**, 2–12.

Jones, M. C. Personality antecedents and correlates of drinking patterns in women. *Journal of Consulting and Clinical Psychology,* 1971, **36**, 61–69.

Keeler, M. H. Motivation for marijuana use: A correlate of adverse reaction. *American Journal of Psychiatry,* 1968, **125**, 386–390.

Keniston, K. *The uncommitted.* New York: Delta, 1965.

Kleber, H. K. Student use of hallucinogens. *Journal of American College Health Association,* 1965, **14**, 109–117.

Klein, J., & Phillips, D. From hard to soft drugs: Temporal and substantive changes in drug usage among gangs in a working class community. *Journal of Health and Social Behavior,* 1968, **9**, 139–145.

Knight, R. C., & Prout, C. T. A study of results in hospital treatment of drug addicts. *American Journal of Psychiatry,* 1951, **108**, 303–308.

Kuehn, J. L. Student drug user and his family. *Journal of College Student Personnel,* 1970, **11**, 404–413.

Laskowitz, D. The adolescent drug addict: An Adlerian view. *Journal of Individual Psychology,* 1961, **17**, 68–79.

Leary, T., Metzmer, R., & Alpert, R. *The psychedelic experience: A manual based upon the Tibetan Book of the Dead.* New Hyde Park, N.Y.: University Books, 1964.

Leeds, D. P. Personality patterns and modes of behavior of male adolescent narcotic addicts and their mothers. *Dissertation Abstracts,* 1965, **26**, 2861.

Lewis, J. M., & Osberg, J. W. Treatment of narcotic addicts: Observations on institutional treatment of character disorder. *American Journal of Orthopsychiatry,* 1958, **28**, 730–749.

Liebert, R. S. Drug use: Symptom disease or adolescent experimentation: The task of therapy. *College Health,* 1967, **16**, 25–29.

Little, R. B., & Pearson, M. M. The management of pathologic interdependence in drug addiction. *American Journal of Psychiatry,* 1966, **123**, 554–560.

Ludwig, A., & Levine, J. Patterns of hallucinogenic drug abuse. *Journal of the American Medical Association,* 1965, **191**, 92–96.

MacKay, J. Clinical observations on adolescent problem drinkers. *Quarterly Journal of Studies on Alcohol,* 1961, **22**, 124–134.

Maddox, G. L. Drinking prior to college. In G. L. Maddox (Ed.), *The domesticated drug: Drinking among collegians.* New Haven, Conn.: College and University Press, 1970.

Malcolm, A. I. Drug abuse and social alienation. *Today's Education,* 1970, **59**, 28–31.

Mamlet, L. N. Consciousness-limiting side effects of consciousness-expanding drugs. *American Journal of Orthopsychiatry,* 1967, **37**, 296–297.

Mason, P. The mother of the addict. *Psychiatric Quarterly Supplement,* 1958, **32**, 189–199.

McAree, C. P., Steffenhagen, R. A., & Zheutlin, L. S. Personality factors in college drug users. *International Journal of Social Psychiatry,* 1969, **15**, 102–106.

McClelland, D. C. The power of positive drinking. *Psychology Today,* 1971, **4**, 40–41, 78–79.

McGlothlin, W. H., & Cohen, S. Use of hallucinogenic drugs among college students. *American Journal of Psychiatry,* 1965, **122**, 572–574.

Mitchell, K. R., Kirkby, R. J., & Mitchell, D. M. Notes on sex differences in student drug usage. *Psychological Reports,* 1970, **27**, 116.

Mizner, G. L., Barter, J. T., & Werme, P. H. Patterns of drug use among college students: A preliminary report. *American Journal of Psychiatry,* 1970, **127**, 15–24.

Mizruchi, E. H., & Perrucci, R. Prescription, proscription and permissiveness: Aspects of norms and deviant drinking behavior. In G. L. Maddox (Ed.), *The domesticated drug: Drinking among collegians.* New Haven, Conn.: College and University Press, 1970.

Modlin, H. C., & Montes, A. Narcotic addiction in physicians. *American Journal of Psychiatry,* 1964, **121**, 358.

Monroe, J. J., Miller, J., & Lyle, W. The extension of psychopathic deviancy scales for the screening of addict patients. *Educational and Psychological Measurement,* 1964, **2**, 47–56.

Murray, J. B. Drug addiction. *Journal of General Psychology,* 1967, **77**, 41–68.

Norton, W. A. The marijuana habit: Some observations of a small group of users. *Canadian Psychiatric Association Journal,* 1968, **13**, 163–173.

Nyswander, M. *The drug addict as a patient.* New York: Grune & Stratton, 1956.

Parker, F. Self-role strain and drinking disposition at a prealcoholic age level. *Journal of Social Psychology,* 1969, **78**, 55–61.

Pearce, J. The role of education in combatting drug abuse. *Journal of School Health,* 1971, **41**, 83–88.

Pearlman, S. Drug use and experience in an urban college population. *American Journal of Orthopsychiatry,* 1968, **38**, 503–514.

Rettig, S., & Pasamanick, B. Subcultural identification of hospitalized male drug addicts: A further examination. *Journal of Nervous and Mental Disease,* 1964, **139**, 83–86.

Robins, L. N. *Deviant children grown up.* Boston: Williams & Wilkins, 1966.

Rogers, E. Group influence on student drinking behavior. In G. L. Maddox (Ed.), *The domesticated drug: Drinking among collegians.* New Haven, Conn.: College and University Press, 1970.

Rosenberg, C. M. Young drug addicts: Addiction and its consequences. *Medical Journal of Australia,* 1968, **1**, 1031–1033.

Rosenberg, C. M. Determinants of psychiatric illness in young people. *British Journal of Psychiatry,* 1969, **115**, 907–915.

Rosenbloom, J. R. Notes on Jewish drug addicts. *Psychological Reports,* 1969, **5**, 769–772.

Sadava, S. W. *College student drug use: A social-psychological study.* (Doctoral dissertation, University of Colorado) Boulder, Colorado: University Microfilms, 1970, No. 70–23749.

Savitt, R. A. Psychoanalytic studies on addiction. *Psychoanalytic Quarterly,* 1963, **32**, 43–57.

Scher, J. Patterns and profiles of addiction and drug abuse. *Archives of General Psychiatry,* 1966, **15**, 539–551.

Skolnick, J. H. Religious affiliation and drinking behavior. *Quarterly Journal of Studies on Alcohol,* 1958, **19**, 452–470.

Smart, R. G., & Fejer, D. Illicit drug users: Their social backgrounds, drug use and psychopathology. *Journal of Health and Social Behavior,* 1969, **10**, 297–308.

Snyder, C. R. *Alcohol and the Jews.* Glencoe, Ill.: The Free Press, 1958.

Sorio, J. Vice as a psychiatric infirmity. *Actas Lusa Espanoles de Neurologia y Psiquiatria,* 1967, **26**, 171–181.

Steffenhagen, R. A., McAree, C. P., & Zheutlin, L. S. Social and academic factors associated with drug use on the University of Vermont campus. *International Journal of Social Psychiatry,* 1969, **15**, 92–96.

Suchman, E. A. The hang-loose ethic and the spirit of drug use. *Journal of Health and Social Behavior,* 1968, 9, 146–155.

Vaillant, G. E. Parent-child cultural disparity and drug addiction. *The Journal of Nervous and Mental Disease,* 1966, **142**, 534–539. (a)

Vaillant, G. E. A twelve-year follow-up of New York narcotic addicts: Some social and psychiatric characteristics. *Archives of General Psychiatry*, 1966, **15**, 599–609. (b)

VOGEL, V., ISBELL, H., & CHAPMAN, K. Present status of narcotic addiction. *Journal of the American Medical Association*, 1948, **138**, 1019–1026.

WAKEFIELD, D. (Ed.) *The addict*. Greenwich, Conn.: Fawcett, 1963.

WELPTON, D. F. Psychodynamics of chronic LSD use. *Journal of Nervous and Mental Disease*, 1968, **147**, 377–385.

WILLIAMS, A. Social drinking, anxiety and depression. *Journal of Personality and Social Psychology*, 1966, **3**, 689–693.

WILLIAMS, A. Psychological needs and social drinking among college students. *Quarterly Journal of Studies on Alcohol*, 1968, **29**, 355–363.

WILLIAMS, A. College problem drinkers: A personality profile. In G. Maddox (Ed.), *The domesticated drug: Drinking among collegians*. New Haven, Conn.: College and University Press, 1970.

ZIMMERING, P. Drug addiction in relation to problems of adolescence. *American Journal of Psychiatry*, 1952, **109**, 272–278.

ZUCKER, R. Sex-role identity patterns and drinking behavior among adolescents. *Quarterly Journal of Studies on Alcohol*, 1968, **22**, 868–884.

(Early publication received July 7, 1972)

Psychological Bulletin
1973, Vol. 79, No. 2, 107–109

PROCESSES OF FEAR REDUCTION IN SYSTEMATIC DESENSITIZATION:

AN ADDENDUM TO WILSON AND DAVISON (1971)

LARRY D. REID[1]

Bradley University

Recent studies are discussed in the context of Wilson and Davison's review of fear reduction techniques tested with animals. In each study, rats had positive intracranial stimulation paired with circumstances eliciting avoidance. Results demonstrate the possibility of counterconditioning and thereby provide an experimental basis for certain procedures of systematic desensitization.

Wilson and Davison (1971), in citing results of Moltz and Nelson, suggested that the primary, and perhaps only, function of a "counterconditioner," such as food with animals and relaxation with human beings, was to act as a facilitator of exposure. They said Nelson's studies "indicate strongly that the provision of a stimulus that supposedly elicits an anxiety-competing response (food, chlorpromazine) facilitates extinction of fear through the increased exposure to the CS which it can bring about [p. 5]." Studies providing strong evidence for the efficiency of counterconditioning in reducing indexes of fear used the conditioned suppression and passive-avoidance paradigms and not the active-avoidance paradigm which may be the better analogue for desensitization. Consequently, Wilson and Davison were hesitant to conclude that these findings were applicable. Other studies purporting to demonstrate the efficacy of counterconditioning were criticized primarily on methodological grounds.

Rats are subjects of choice for testing many propositions; however, it is difficult to manipulate affective states in rats as relaxation or pleasant imagery allows in desensitization therapy. The trained client can regulate, at request, sequencing of anxiety-eliciting procedures and relaxation rather precisely. With animals, providing food has been the most frequently used counterconditioner. However, some levels of fear inhibit feeding (Estes & Skinner, 1941) making it difficult to balance propensity to eat against fear so that food can countercondition fear. Also, it is not known when food has its effect. New technology, however, now allows precise control of intensity and temporality of affective states in rats through the procedures of direct electrical stimulation of hypothalamus. Extensive research with intracranial stimulation (ICS) has made it possible to impose upon rats an unambiguously positively reinforcing stimulus (Olds, 1962; Reid, 1967). Furthermore, ICS can mask or attenuate pain and fear (Reid, 1967). Three recent studies have capitalized on the advantages provided by ICS and have tested the efficacy of counterconditioning while controlling important variables such as exposure time of the conditioned stimulus (CS) and temporality of counterconditioner.

Gordon and Baum (1971) trained rats to avoid footshock by jumping to a retractable platform. Prior to training for avoidance, rats had been fixed with electrodes to deliver positive ICS. After avoidance was established, footshock was terminated and rats were prevented from jumping by removal of the platform (response prevention). One group also received trains of positive ICS throughout the prevention time. The group that received response prevention plus ICS was most efficiently deconditioned (median measures of response strength were less by a factor of 25).

Monico (1971) fixed rats with electrodes to stimulate positive hypothalamic areas. They were then placed in a chamber delivering footshock 5 seconds after light (CS) onset with CS and shock continuing until a lever press. If there were a press before 5 seconds of CS, footshock was avoided. At first, rats were shaped to avoid, that is, CS was turned off and shock avoided for successive approximations to avoidance. Shaping disrupted the freezing characteristic of rats in this paradigm and led to reasonably efficient avoidance and efficient escape when avoidance was missed. After shaping, rats were given 141 trials before being divided into six groups to receive one of six treatments. During all treatments the lever was removed. Treatments conformed to a 2 × 3 factorial design with ICS or no ICS as one factor. The other factor was CS used during training, CS of gradually increasing intensity from a faint intensity to CS used in initial training, and no CS. The group that received full-intensity CS (15 presentations of 5 seconds each) was a response prevention group. The group that received full-intensity CS with ICS (.5 seconds trains, 1 and 3 seconds after CS onset) was a counterconditioning group. After treatments, the lever was returned and CS presented 220 times during two sessions as in training but without footshock. On all measures of response perseveration, (avoidance responses, total lever presses, and time CS remained on) deconditioning was twice as effective with CS plus ICS treatment than with any other treatment.

---

[1] Requests for reprints should be sent to Larry D. Reid, Department of Psychology, Bradley University, Peoria, Illinois 61606.

Stone (1971) trained rats as Monico did up to the time of treatments. As therapy, one group had the same 5-second CS as used in training (response prevention). The other three groups had the same CS but had ICS either before, during, or after CS. The lever was returned and the rats were given 220 trials during which CS came on but footshock did not. On all measures, rats having ICS during and after CS were more efficiently deconditioned than rats with no ICS or ICS just before CS. There were insignificant differences between groups receiving ICS during and after CS and between groups with no ICS and ICS before CS.

It is clear that different experimenters in different laboratories using different response measures but similar treatments obtained the same result: Changing affect with positive ICS during circumstances, previously paired with an aversive event, efficiently attenuated persisting avoidance. Conditioned stimulus exposure time was the same regardless of whether or not rats received ICS; consequently, ICS did not merely facilitate exposure to the CS. We need not hedge; there is clear evidence from the animal laboratory that extended exposure to fear-eliciting conditions without aversive consequences typically leads to deconditioning of the conditioned emotional response, of passive avoidance, and of active avoidance, but that deconditioning in each paradigm is facilitated when CS is accompanied by a positive affective stimulus. In addition to being a facilitator of exposure, an anxiety-competing or positive stimulus can be a counterconditioner.

The mechanism of counterconditioning is unknown. A start toward specifying the mechanism seems to be embodied in those attempts to specify the interrelatedness of forebrain systems mediating positive and negative affect (Stein, 1964; Young, 1967). The structure of the limbic system, allowing reverberation and positive feedback, is an adaptation for maintaining behavior in the absence of immediate stimulus control. A counterconditioner may be any event that would interrupt or diminish reverberation maintaining fear.

Studies reviewed here lead to a different emphasis from that of Wilson and Davison (1971). But more importantly, they are a demonstration of a general technique that will allow for preclinical screening of potential therapies. Using avoidance paradigms with animals and ICS, it is possible to vary systematically parameters of strength of avoidance and parameters of "therapy" as spacing, intensity, quality, and frequency of affective change. Such parametric studies should yield conclusions about the most efficient way to reduce fear and anxiety without harmful side effects.

REFERENCES

Estes, W. K., & Skinner, B. F. Some quantitative properties of anxiety. *Journal of Experimental Psychology*, 1941, **29**, 390–400.

Gordon, A., & Baum, M. Increased efficacy of flooding (response prevention) in rats through positive intracranial stimulation. *Journal of Comparative and Physiological Psychology*, 1971, **75**, 68–72.

Monico, K. F. Efficient extinction of persistent avoidance responding. Unpublished master's thesis, Bradley University, 1971.

OLDS, J. Hypothalamic substrates of reward. *Physiological Reviews,* 1962, **42,** 554–604.

REID, L. D. Reinforcement from direct electrical stimulation of the brain. Unpublished doctoral dissertation, University of Utah, 1967.

STEIN, L. Reciprocal action of reward and punishment mechanisms. In R. G. Heath (Ed.), *The role of pleasure in behavior.* New York: Hoeber, 1964.

STONE, D. F. Models for reduction of fear and avoidance. Unpublished master's thesis, Bradley University, 1971.

WILSON, E. T., & DAVISON, G. C. Processes of fear reduction in systematic desensitization: Animal studies. *Psychological Bulletin,* 1971, **76,** 1–14.

YOUNG, P. T. Affective arousal: Some implications. *American Psychologist,* 1967, **22,** 32–40.

(Received July 28, 1971)

since 48 of the 81 small squares are under the OD curve.

Performing Standard Significance Tests with an Ordinal Dominance Curve

Most or all of the computation in at least five common nonparametric significance tests can be avoided once an OD curve is constructed. The five are the 2 × 2 chi-square test, the median test, the Mann-Whitney *U* test, the Wald-Wolfowitz runs test, and the two-sample Kolmogorov-Smirnov test.

Consider first the 2 × 2 chi-square test computed when a worker divides a scale at some point and tests the null hypothesis that equal proportions of the two groups lie above (or below) that point. If the groups are roughly equal in size, then the phi or correlation for any point can be read by finding the appropriate values of P_1 and P_2 in Figure 3.[4] For example, in Figure 2 the point between scale values of 6 and 7 corresponds to a phi of about .28. If the conditions for a chi-square test are met, then the value of phi corresponding to a given significance level can be found either from a table for the significance of the Pearson *r* with $N_1 + N_2$ cases, or from the relation $\phi = \sqrt{\chi_1^2/(N_1 + N_2)}$. In the present example, either method shows the critical value at the 5% level to be .37. If sample sizes are too small for this method, Fisher's exact 2 × 2 test may be used instead.

The median test (Conover, 1971, pp. 167–172; Siegel, 1956, pp. 111–116) consists simply of dichotomizing at the median and then performing a 2 × 2 chi-square test as above. The median is readily found in an OD curve, so this test too can be done with little or no computation.

The Mann-Whitney test (Conover, 1971, pp. 224–236; Siegel, 1956, pp. 127–136) also has a remarkably simple relation to an OD curve. The number of small squares below the OD curve equals the number of units by which the rank sum for Group 1 exceeds its minimum possible value of $N_1(N_1 + 1)/2$, and the number above equals the number of units by which the rank sum for Group 2 exceeds its minimum possible value of $N_2(N_2 + 1)/2$. In Figure 1, for example where $N_1 = N_2 = 9$, the minimum possible value of either rank sum is 45. Since there are 33 small squares above the OD curve, the rank sum for Group 2 equals 45 + 33, or 78. From any standard table, this corresponds to a one-tailed probability level of .273. Alternatively, *U* in the Mann-Whitney test is the smaller of the number of squares above or below the OD curve—33 in Figure 1.

A Wald-Wolfowitz runs test (Conover, 1971, pp. 350–356; Siegel, 1956, pp. 136–145) can be performed, once an OD curve has been constructed, simply by counting the segments of the curve. In Figure 1, for example, there are nine vertical or horizontal segments of the OD curve, hence nine runs. With $N_1 = N_2 = 9$, six or fewer runs are required for significance.

The test statistic in the Kolmogorov-Smirnov two-sided two-sample test (Conover, 1971, pp. 309–314; Siegel, 1956, pp. 127–136) is the largest difference between the two cumulative frequency distributions, which equals the largest vertical distance between the OD curve and the diagonal line. In Figure 1, for example, the largest difference is 4/9 or .44; a difference of .67 would be required for significance.

MIXED AND COMPLETE DOMINANCE: NEW VERSIONS OF HYPOTHESES ABOUT DISPERSION AND ELEVATION

No standard statistical test examines hypotheses of dispersion or elevation in the form in which they are frequently most interesting to the behavioral scientist.

---

[4] Figure 3 was drawn with the assumption that the two groups are exactly the same size. However, investigation has shown that it is quite accurate within the range of figures usually encountered. For example, if the larger group is not more than 5/4 the size of the smaller, and if both within-group proportions are between .1 and .9, and if phi as read from Figure 3 is .30, then it can be shown that the true phi is between .29 and .31. The error is thus so small that it can be largely ignored. In general, the margin of error is greatest around the edges of Figure 2, and smallest in the center. Since the major reason for computing phi is usually to test the significance of its difference from zero, and since the chi-square significance test is also usually inaccurate at the extreme edges of Figure 2, we can conclude that moderate inequalities of sample sizes place few or no more limitations on the use of Figure 3 than are placed by the limitations of the chi-square test.

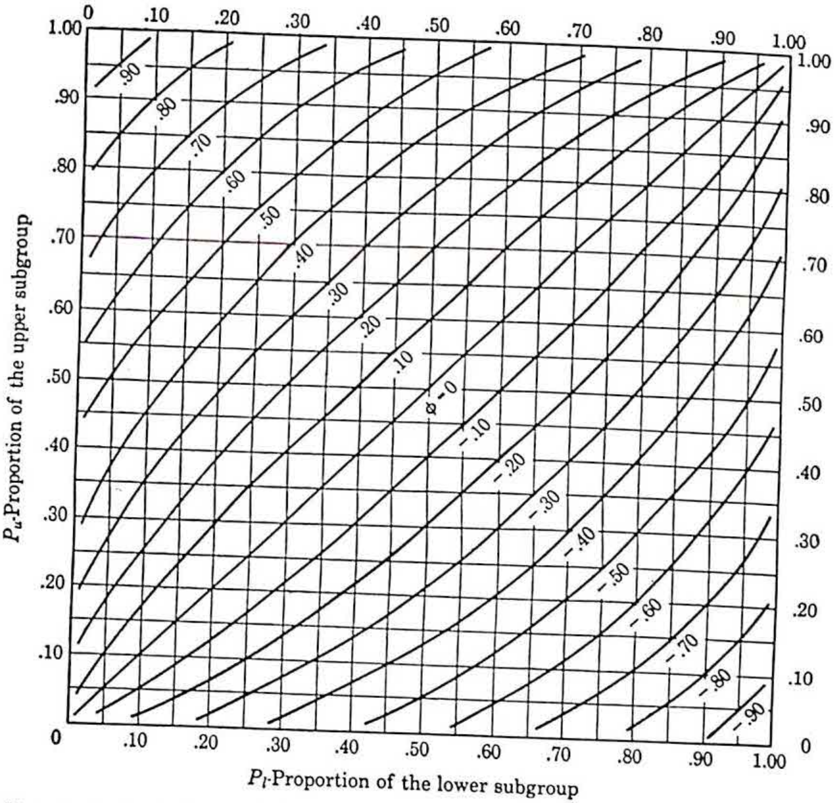


FIG. 3. A chart for making graphic estimates of phi (ϕ) when two groups are the same size. (Reprinted with permission of the McGraw-Hill Book Co. From J. P. Guilford, *Psychometric methods.* New York: McGraw-Hill, 1954.)

Mixed Dominance and Dispersion

A typical example of a hypothesis about dispersion is given by Siegel (1956): "Women who have personal difficulty in handling aggressive impulses are *more extreme* than others in their judgments of hostility in film characters—some underattribute and others overattribute [p. 149]." The data used to test this hypothesis were the hostility-attribution scores of a group of women with difficulties controlling their aggressive impulses, and another group of women without such difficulties. The data are given in Figure 1 of the present paper. The hypothesis predicts that scores in Group 1 should be more variable than those in Group 2. This prediction was confirmed by the Siegel-Tukey (1960) test and the Freund-Ansari (1957) test, so the hypothesis quoted above would appear to be confirmed. Thus a double conclusion is accepted: Some members of Group 1 overattribute and some underattribute. Inspection of Figure 1, however, lends little support to the conclusion that some underattribute. The two distributions are nearly identical at the low end of the scale, since the OD curve roughly follows the diagonal. As we shall show more fully later, the data are consistent with the hypothesis that the two group distributions are identical below their medians but that Distribution 1 has a longer tail at the high end. Thus the conclusion that the two distributions differ in variability is supported, but the original hypothesis quoted from Siegel is not. This is unfortunate, since the quoted hypothesis is clearly more interesting than the hypothesis actually confirmed by the Siegel-Tukey and Freund-Ansari tests.

The quoted hypothesis states that both tails of Distribution 1 contain more cases than the tails of Distribution 2; that is, $P_1 - P_2$ is positive at the low end of the scale and negative at the high end. Thus the OD curve must cross the diagonal, being below it at the low end and above it at the high

Psychological Bulletin
1973, Vol. 79, No. 2, 110-116

COMPARING TWO GROUPS BY SIMPLE GRAPHS[1]

RICHARD B. DARLINGTON[2]

Cornell University

Standard significance tests for testing whether two independent groups are equal in elevation or dispersion do not test the forms of those hypotheses which are frequently of most interest to behavioral scientists. More interesting forms of those hypotheses can be tested easily and quickly by drawing a simple graph. The same graph can also be used to perform several standard tests. In all these tests, the procedure is so simple that the chance of computational error is much smaller than with standard procedures.

When a significance test is used to reject the null hypothesis that two independent sets of scores are identically distributed, the alternative hypothesis is almost always one of three types. In order of their frequency of use they are: hypotheses specifying differences in elevation or location (usually tested by the t test, Mann-Whitney U test, or the median test or other 2×2 tests); hypotheses specifying differences in variation or dispersion (usually tested by the F test on variances or by the nonparametric Siegel-Tukey or Freund-Ansari tests); and hypotheses which do not specify the nature of the differences (usually tested by the Wald-Wolfowitz runs test or the Kolmogorov-Smirnov test). A simple graph makes it possible to perform most of these tests in ways which are quick, easy to understand, and almost completely free of the chance of computational error. Some new tests, which test hypotheses about elevation and dispersion which are frequently more interesting to behavioral scientists than any of the specific hypotheses tested by standard tests, are also performed easily. The graph will be called an *ordinal dominance* or OD curve, though a somewhat similar curve is called a P-P curve by Wilk and Gnanadesikan (1968).

ORDINAL DOMINANCE CURVES AND THEIR RELATION TO SOME FAMILIAR STATISTICAL PROCEDURES

Construction of an OD curve is very simple, as can be illustrated by Figure 1. First, a rectangle is drawn with length and height, respectively, of N_1 and N_2, the sizes of the two groups. Then the scores of Group 1 are ordered and written immediately below the rectangle. The scores of Group 2 are ordered and written immediately to the left of the rectangle.[3]

Each segment of an OD curve represents the score broadside to it. Starting at the lower left corner, the OD curve is drawn by drawing a continuous curve upward and to the right so that each successive curve segment represents the smallest score not represented by a previous segment. In Figure 1, for example, the smallest score is 0 from Group 1, so the first segment of the OD curve is horizontal. The next smallest score is 3 from Group 2, so the next segment is vertical; and so on. Thus the $N_1 + N_2$ scores are represented in the order of their size, from smallest to largest, as shown in Figure 1.

It is often convenient to write the within-group rank orders adjacent to the scores, and the values of P_1 and P_2, the corresponding proportions, on the right and upper sides of the figure. Each value of P is written adjacent to a line, not a square, of the grid. Thus a P value of, say, .5 corresponds to the *division between* the two median scores of a group of even size, rather than to one score or the other. A straight line from the rectangle's lower left to upper right corner may also be drawn showing the points where $P_1 = P_2$.

---

[1] For many helpful criticisms and suggestions the author is indebted to his students Paul Cieslak, Jean Rom, Sharon Weinberg, Robert Wunderle and others, and to his colleagues Jason Millman, T. A. Ryan, and Gordon Wilcox. Any errors remain his own.

[2] Requests for reprints should be sent to Richard B. Darlington, Department of Psychology, Uris Hall, Cornell University, Ithaca, New York 14850.

[3] These data are taken from Siegel (1956, p. 149).

When scores from different groups are tied, the OD curve is drawn diagonally across the spaces representing the tied scores. This is illustrated in Figure 2, which shows the ratings on a 9-point scale given by subjects in an experiment by Jones and Panitch (1971). The topmost segment of the OD curve, for example, represents all the scores of 9—four scores from Group 1 and one score from Group 2. It is drawn diagonally across the space represented by those scores.

An OD curve is a type of cumulative frequency curve; it shows the cumulative frequency of Group 2 as a function of the cumulative frequency of Group 1.

The shape of the OD curve is determined entirely by the rank order of the observations, and that rank order can be reconstructed completely from the shape. The curve is thus not a *summary* of the data; rather, it is a *complete representation* of the rank-order properties of the data.

Basic Descriptive Statistics Easily Read from an Ordinal Dominance Curve

Four common types of descriptive statistics are easily read from an OD curve. None of these is easily read from standard frequency curves.

The slope of the straight line between any two points on an OD curve equals the relative frequency of Groups 1 and 2 within the interval bounded by those points. For example, consider the point on the OD curve in Figure 2 between segments representing scale values of 5 and 6. A straight line from the lower left corner to this point has a slope of 6/3, or 2, so twice as many subjects from Group 2 as from Group 1 are in the interval 1–5.

The vertical distance by which the OD curve lies above the straight diagonal line at any point equals $P_2 - P_1$ at that point.

If any point in an OD curve is connected with the diagonal by a straight line with a slope of -1, the value of P_1 or P_2 where this line meets the diagonal equals the proportion of the combined group falling below the point. For example, a line through the midpoint of the diagonal in Figure 2, with slope -1, intersects the OD curve between segments representing scores of 6 and 7. Thus 50% of the combined group falls below this point.

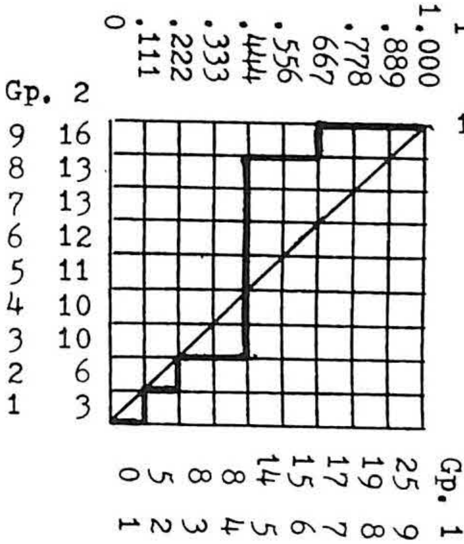


FIG. 1. An OD curve for data with no ties between groups.

Each small square represents a comparison between the Group 1 member below the square and the Group 2 member to its left. Since the squares below the OD curve represent comparisons in which the Group 1 member is higher than the Group 2 member, the proportion of the rectangle's area under an OD curve equals the probability that a randomly chosen member of Group 1 will score higher on the scale than a randomly chosen member of Group 2 (with any ties split evenly between the groups). For example, in Figure 1 this probability is 48/81 or .59,

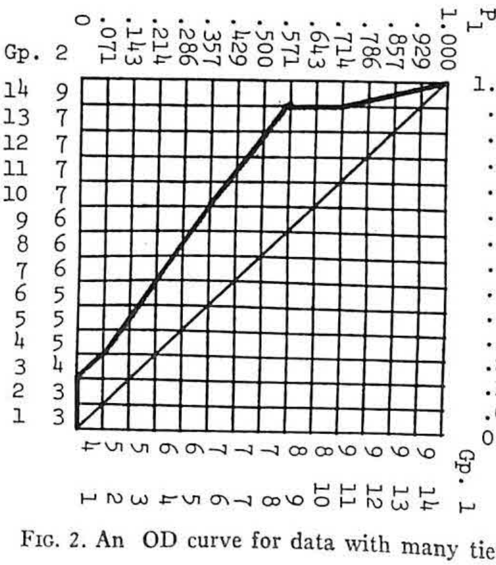


FIG. 2. An OD curve for data with many ties.

end. We shall define the case in which the OD curve touches or crosses the diagonal as the case of *mixed dominance*. The opposite case, in which the OD curve stays entirely on one side of the diagonal, will be termed *complete dominance*. The Siegel hypothesis is, then, the hypothesis that mixed dominance occurs in the population studied.

This is not an isolated example. In many applications a difference in dispersion is of little interest unless it implies mixed dominance. Even if a difference in variances is interesting, a further demonstration of mixed dominance usually makes it still more interesting.

Complete Dominance and Elevation

Hypotheses about differences in elevation are usually tested by the *t* test, the Mann-Whitney rank-sum test, or the median test. One group may be "higher" than the other in the sense tested by any of these tests (or all three together) even though mixed dominance exists. Yet, if mixed dominance exists, then a general, unqualified statement like "Group 2 is higher than Group 1" seems questionable, since the proportion of Group 1 above some scale points then exceeds that of Group 2. For some purposes, then, Group 1 may be "better" or "higher" than Group 2, even though Group 2 is higher in the sense tested by any standard test.

Even if a measure is clearly interval or ratio, like number of right answers on a multiple-choice test, the group with the lower true mean may have a higher proportion of members above some important point on the scale, due to differences in shape or variance of within-group distributions. This result cannot occur under an assumption sometimes mentioned by statisticians—that two distributions are identical in variance and shape (Conover, 1971, p. 225). However, the assumption is so narrow that it is almost always ignored in practice (as in Conover, 1971, p. 226), and is not even defined for ordinal scales.

In many applications, then, the hypotheses supported by standard tests of elevation are of much less interest than the hypothesis of complete dominance. If Group 2 completely dominates Group 1, then the proportion of Group 2 exceeds that of Group 1 above any point on the scale. Group 2 is then also higher than Group 1 in the sense tested by the Mann-Whitney test; that is, a randomly chosen member of Group 2 has a better-than-even chance of being higher than a randomly chosen member of Group 1. If the scale is interval, the mean of Group 2 is not only higher than that of Group 1 but would remain higher after any monotonic transformation of the scale. This is of particular interest to a worker who thinks of his observed scale as being monotonically, but not necessarily linearly, related to some other variable of more fundamental interest. For example, a worker may think of an exam score as having a monotonic but nonlinear relation to "ability" or "knowledge," or a decision theorist may think of an observed scale as having a nonlinear relation to "utility." Complete dominance of Group 2 would then imply that Group 2 is higher than Group 1 in mean ability or knowledge or utility, even though the precise relations between these variables and the observed measures are not specified. Thus, if complete dominance is demonstrated, there need be no hesitation in stating a general unqualified conclusion like "Group 2 is higher than Group 1."

Procedures for Demonstrating Mixed and Complete Dominance

A worker should first attempt to reject the identity hypothesis (the hypothesis that the two group distributions are identical in every way), using one of the tests mentioned above. If the identity hypothesis is rejected, and the sample OD curve crosses the diagonal line, the next step is to attempt to reject the complete dominance hypothesis, thereby accepting the mixed dominance hypothesis as the only remaining alternative. To do this, the worker should divide his sample into two subsamples, one consisting of those subjects above the point where the OD curve crosses the diagonal line, and the other consisting of those below that point.[5] The worker then should perform a test for differences in elevation in each subsample, attempting to show that the OD curve goes significantly above the diagonal line in one subsample and significantly below it in the other.[6] The mixed dominance hypothesis then is accepted at the higher (i.e., less significant) of these two probability levels. In Figure 1, for example, the upper and lower subsamples contain, respectively, five and four subjects from each group. In the upper subsample, the Mann-Whitney test rejects the identity hypothesis at the .016 level (one-tailed). In the lower subsample, however, the comparable probability level is merely .243. The hypothesis of mixed dominance thus is not accepted in this example.

---

[5] If the OD curve crosses the line at several points, any one of these may be chosen. We suggest, for reasons concerning the power of the tests to follow, choosing the point closest to the median of the total sample.

If the identity hypothesis is rejected but mixed dominance is not shown, a worker may wish to support a conclusion of complete dominance. The simplest way to reach such a conclusion is to simply assume from the nature of the problem that mixed dominance does not exist. Since the identity hypothesis has been rejected, the complete dominance hypothesis then is accepted as the only remaining alternative. Though it may appear trivial at this point in the exposition, this conclusion answers a question of considerable interest to behavioral scientists. A worker who has found a significant difference in the elevation between two groups by a *t* test or other standard test is often dissatisfied with the narrowness of the conclusion that he is technically allowed to draw from this result. He often asks whether there is some assumption which, if made, will make his result of more general interest. We have already seen that an assumption often mentioned by statisticians, that two group distributions are identical except for their means, is inapplicable to ordinal scales and too restrictive when applied to interval scales. We now see that a much less restrictive assumption, the assumption of no mixed dominance, enables a worker to draw a very general conclusion, the conclusion of complete dominance, from a *t* test or other standard test of elevation.

A worker who is unwilling to assume absence of mixed dominance may still reach a conclusion of considerable interest by a further analysis of his data. We saw earlier how he can perform a 2 × 2 chi-square-type test, by comparing a point on an OD curve to Figure 3, after determining the value of phi necessary for significance. In a few seconds, a worker can perform this test for every point on the OD curve, marking the points which turn out to be significant. The points very near the two ends of the OD curve can never be significant by this test, since the OD curve necessarily approaches the diagonal at its two ends. However, there may be a substantial range in the middle of the curve in which every point is significant. If so, we can say that a complete dominance relation exists within this range.[7] This implies that the proportion of, say, Group 2 exceeds the proportion of Group 1 above every point within the range, and that the mean of Group 2 exceeds that of Group 1 for any monotonic scale transformation which is flat outside the range.

This procedure has the strong advantage over *t* tests or other standard techniques that it clearly specifies the range over which a conclusion about a difference in elevation is known to be valid. The common-sense observation that conclusions about geniuses or star athletes should not be drawn from random samples of schoolboys is thus made more specific. A worker who uses a standard test on elevation to draw an unrestricted conclusion like "farm boys in general are more physically fit than town boys" (as in Conover, 1971, p. 226) will perhaps tempt some readers to extrapolate the conclusion beyond the scale levels actually warranted by the data. Even a careful and well-intentioned reader will be unclear about just how much extrapolation is warranted. Using the present technique, however, the investigator will be able to make a precise statement about the range over which his conclusion is known to apply.

---

[6] The fundamental fact which makes this procedure acceptable is the fact that if the null hypothesis is correct, each of the Mann-Whitney statistics will be distributed as shown in the Mann-Whitney tables in the subset of samples in which the OD curve touches the 45-degree line at the point chosen. In other words, this significance test is based not on all possible samples, as most significance tests are, but on all samples with a particular characteristic. This technique is the basis of the Fisher exact 2 × 2 test and other randomization tests. In the present case, the two Mann-Whitney tests are statistically independent of each other in the subset of samples described.

[7] To protect against the possibility that a range of complete dominance may be found just by chance, due to the large number of possible ranges, a simple and useful rule is to require that any such range must include the sample median. Within such a range, the usual proscription against multiple significance tests does not apply, since it is required that *all* tests, not just one of many, be significant.

In summary, the present technique has the following advantages over conventional techniques: (*a*) It specifies precisely the range over which its conclusions are know to apply; (*b*) its conclusions within that range are very strong and interesting; (*c*) these conclusions are reached without any assumptions beyond ordinal scaling of variables; (*d*) the procedures used are quick, easy to understand, and free of serious chance of computational error.

REFERENCES

CONOVER, W. J. *Practical nonparametric statistics.* New York: Wiley, 1971.

FREUND, J. E., & ANSARI, A. R. *Two-way rank sum test for variances.* (Technical Report No. 34) Blacksburg: Virginia Polytechnic Institute, 1957.

GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1954.

JONES, S. C., & PANITCH, D. The self-fulfilling prophecy and interpersonal attraction. *Journal of Experimental Social Psychology,* 1971, **3**, 356–366.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

SIEGEL, S., & TUKEY, J. W. A nonparametric sum of ranks procedure for relative spread in unpaired samples. *Journal of the American Statistical Association,* 1960, **55**, 429–444.

WILK, M. B., & GNANADESIKAN, R. Probability plotting methods for the analysis of data. *Biometrika,* 1968, **55**, 1–17.

(Received July 29, 1971)

Psychological Bulletin
1973, Vol. 79, No. 2, 117–126

ON THE FUNCTIONS OF STRUCTURAL PARAPHRASE: THE VIEW FROM THE PASSIVE VOICE

MOSHE ANISFELD[1] AND IRENE KLENBORT

Yeshiva University

An integration of concepts from traditional and generative linguistics and a reinterpretation of experimental research in psychology have led to the following conclusions: (*a*) Chomskyan deep structure captures the semantic commonality between the active voice and the passive voice. (*b*) The semantic differences between the voices are attributable primarily to properties of the passive, the active serving as a neutral base by reference to which the passive takes on its significance. This asymmetrical relation between the voices is a reflection of the marked status of the passive vis-a-vis the unmarked active. (*c*) The functions distinguishing the passive from the active are (*i*) to allow for emphasis of the logical object and to provide discourse continuity and (*ii*) to enable the speaker to build a presuppositional infrastructure into his message. For instance, the sentence *The boy was bitten by the dog* presupposes that the boy was bitten by someone and asserts that that someone was the dog. Actives are considered to be generally presuppositionally neutral.

Language affords a rich variety of opportunities for changing the syntactic structure of a message while retaining its essential content and lexical composition. For instance, Sentences 1–5 below exemplify only a sample of the structural paraphrases of the content involved.

1. The police believe that John has stolen the jewel.
2. The police believe that the jewel has been stolen by John.
3. John is believed by the police to have stolen the jewel.
4. The jewel is believed by the police to have been stolen by John.
5. What the police believe John to have stolen is the jewel.

What are the functions of structural paraphrase? This paper attempts to show that some recent work in generative linguistics involving, in particular, the notions *presupposition* and *markedness* point to new avenues for answering this question. The discussion will be developed around the active-voice–passive-voice relation, which has generally been considered as the paradigmatic case of structural paraphrase.

The paper is divided into four sections. The first section deals with the formulation of the essential commonalities between the two voices. The second section deals with differences between them in terms of the notion of markedness. The third section discusses the semantic differences between the active and the passive as they have been traditionally conceptualized by linguists and studied experimentally by psychologists. The studies reviewed in the third section are reinterpreted from the perspective of markedness developed in the preceding section. The fourth section elaborates the presuppositional differences between the two voices and examines the interaction of presupposition and markedness.

Active–Passive Commonalities

The active voice and the passive voice are generally recognized as basically synonymous. Thus, for instance, Sentences 6 and 7 express the same content.

6. The dog bit the boy.
7. The boy was bitten by the dog.

These two sentences are identical with respect to truth value: if sentence 6 is true, Sentence 7 is also true, and vice versa. However, the close affinity between the two voices had not received formal expression till the advent of generative transformational grammar (Chomsky, 1957). The introduction of a deep level of syntactic analysis in addition to the surface level of previous theory provided transformational grammar the vehicle for formulating the essential similarity between the two voices. On the surface level, the simple active (as in Sentence 6 above) has the structure shown in 8 below and the passive has the structure shown in 9.

---

[1] Requests for reprints should be sent to Moshe Anisfeld, Department of Psychology, Yeshiva University, New York, New York 10003.

8. NP_1 — Aux — V — NP_2
9. NP_2 — Aux + be + en — V — by + NP_1

Where: NP = noun phrase (e.g., *the dog, the boy*), V = verb (e.g., *bite, break*), Aux = forms that specify number and tense of verb, en = participial form of verbs (e.g., *bitten, broken*). Comparison of Structures 8 and 9 shows that what is the grammatical subject (NP_1) in the active is the grammatical object in the passive and what is the grammatical subject in the passive (NP_2) is the grammatical object in the active. On the deep-structure level, however, the two constructions have the same representation. The deep structure expresses the logicosemantic relations, which are common to both voices. In deep structure, NP_1 is the logical subject of both Sentences 6 and 7 and NP_2 is the logical object of both sentences. Thus, in the active there is a coincidence between logicosemantic relations and grammatical relations, whereas in the passive, the logical subject is the grammatical object and the logical object is the grammatical subject. The mechanism effecting this inverted mapping of the deep structure onto the surface structure is the passive transformation. This transformation, along with such other transformations as negation and question, was initially considered to be optional (Chomsky, 1957, pp. 42–43, 77–83, 112), but later, in order to prevent transformations from changing meaning, it, along with the other ones, was formulated as obligatory and dependent on the presence of an optional marker in the deep structure (Chomsky, 1965, pp. 103–106, 223; Katz & Postal, 1964, pp. 72–74). This change is of little psychological consequence except as it reflects the restriction of the semantic analysis of generative grammar to deep structure relations and the consequent inability of the theory to accommodate potential differences between the active and the passive. However, the main thrust of generative grammar in its first decade was to account for linguistic *similarities* unrecognized or unexpressed in descriptive theories of language, and this purpose, as we have seen, it achieved with respect to the active and passive constructions. Generative theory also provided a formal justification for the general practice of grammarians to consider the active as more basic than the passive. The greater complexity of the passive is due to the transformation that interchanges the positions of the two noun phrases.

Active–Passive Formal Differences: On Markedness

The greater transformational complexity of the passive in relation to the active can be profitably conceptualized within the theory of markedness. The notion of markedness is designed to distinguish between the typical or normal cases (unmarked) and the ones that in comparison are less typical (marked). The concept was developed in the 1930s within the Prague school of linguistics and has been revived in Greenberg's (1966) work on language universals as well as in generative phonology (Chomsky & Halle, 1968, Ch. 9; Postal, 1968, Ch. 8). In this section, we will first discuss markedness in the realms of phonology, morphology, and semantics, where it has been mostly studied, and then, extend it to our present syntactic case of active-voice–passive-voice relations.

The notion of markedness developed from an observation of asymmetries in the behavior of closely related items. One of the chief manifestations of the asymmetry is the generic behavior of one of the items (unmarked) in comparison with the specific behavior of its mate (marked). Thus, in antonymous adjectival pairs such as *long–short*, the first term stands not only for one pole of the dimension but also for the entire category. This asymmetry is reflected in the status of questions. The question *How long is the paper?* does not imply that it is either long or short. Whereas the question *How short is the paper?* does imply that it is short. Also, *length*, derived from long, refers to the entire dimension. In phonology, the generic character of unmarked forms is reflected in the phenomenon of neutralization. In certain phonological environments only one member of a phonemic pair occurs, representing the entire superordinate archiphoneme. A second characteristic of unmarked items is their relative structural simplicity. Comparing the singular (unmarked) with the plural (marked) noun, Greenberg (1966) has noted that "The singular frequently has no overt mark while the plural is marked by an affix as in English . . . [p. 73]." A third point noted about unmarked forms is that they tend to be distributionally freer than their marked mates. For instance, in English /s/ can precede vowels, consonants, liquids (e.g., sip, spit, ski, sleep) and one semivowel (e.g., sweep), whereas /θ/ is restricted to vowels and the two semi-vowels (e.g., thin, throw, thwart). In phonological analysis, these two sounds /s/ and /θ/ constitute a minimal pair differing only in one feature (stridency). The sound /s/ would be considered unmarked and /θ/ marked. This decision is supported by the next two criteria for markedness. The fourth point concerns greater frequency, within a given language, of the unmarked item. Thus /s/ occurs 12 times as frequently in English as /θ/ (Miller, 1951, p. 86). A fifth characteristic of unmarked items is their wider presence in the languages of the world. In this connection, Chomsky and Halle (1968) have remarked that the least marked consonants in English, including /p t k s n/, "are rarely absent in the phonological system of a language [p. 413]." Finally, unmarked forms tend to interact more closely with other grammatical categories than marked forms. For instance, in German and other gender languages, the adjective agrees in gender with the singular (unmarked) noun but only one adjectival form exists for the plural (marked) noun.

The coordination of these different earmarks of markedness with one another and across the different levels of language (phonological, morphological, syntactic, and semantic) has not been settled yet on the rigorous theoretical level, but their interrelatedness is quite apparent on the intuitive level (see Greenberg, 1966). All the criteria converge on the generalization that the unmarked term is more natural and basic than the marked one. Generative linguists have assigned central importance to the criterion of structural simplicity, in extended form, in the determination of markedness. The extension has been beyond local simplicity of one item in relation to another and involves consideration of the effects on the entire system of generalizations of choosing a particular item as the unmarked one. The notion of simplicity has been treated in formal terms, but it is quite likely that structural simplicity will be correlated with psychological simplicity in internal processing and in production and perception. Discussing markedness in phonology, Postal (1968) has stated that "Ultimately, perhaps some of the strongest evidence for assignment of Marked or Unmarked status will come from physiological and perceptual investigations [pp. 170–171]." Following Roman Jakobson, he has also suggested (Postal, 1968, p. 172) that unmarked items should appear before their marked counterparts in the course of linguistic development in children.

Psychological research, concentrating mainly in the lexical domain, has explored some of the behavioral correlates of markedness. The research of Clark and his associates (Clark, 1969; Clark & Card, 1969, and references therein) has suggested that marked adjectives are coded in memory in terms of their unmarked counterparts. For instance, marked adjectives tended to be recalled much more often as their unmarked counterparts than vice versa. Hamilton and Deese (1971) found that unmarked adjectives, even when their referential meaning was neutral, were classified as more positive than their marked counterparts. Since, however, they have shown that unmarked adjectives are used more frequently, it is not clear whether the differences in evaluation are due directly to markedness or are mediated by frequency. Experiential frequency has been found to be correlated with evaluative judgments (see Anisfeld, 1968; Johnson, Thomson, & Frincke, 1960).

Returning now to the passive from the perspective of the foregoing discussion, it seems clear that the English passive (as well as the passive in other languages, see Greenberg, 1966, p. 86) should be considered marked and the active unmarked. In the first place, the passive is longer and transformationally more complex than the active. The passive entails all the transformations the active does plus the one that interchanges the positions of the two noun phrases. Chomsky (1957, pp. 79–81) has shown that reversing this state of affairs by making the passive closer to the underlying structure and deriving the active by a special transformation would render the grammar more complicated and cumbersome. The passive is also much less frequently used than the active, possibly because of its greater complexity. In an extensive survey of present-day English sentences totalling 320,000 words, Svartvik (1966, p. 46) found that the frequency of passives was one-seventh that of actives. Also, the formation of passives appears to be restricted. As Allen (1959) has noted, "most sentences that are good in the active voice are just grotesque curiosities when put into the passive voice [p. 290]." More strictly, there are some verbs that just do not take passive forms. Excluded are not only intransitive verbs (e.g., sleep), which do not have direct objects, but also some other verbs (e.g., cost, resemble), which have direct objects (see Long, 1961, p. 119). Thus, Sentences 11 and 13 cannot be formed from Sentences 10 and 12, respectively.

10. The book costs ten dollars.
11. *Ten dollars are cost by the book.
12. John resembles President Nixon.
13. *President Nixon is resembled by John.

Lees (1960, p. 8) has observed that verbs that do not take the passive do not take manner adverbials freely. On the basis of this observation, Chomsky (1965, pp. 101–106) has suggested that the passive marker in the deep structure be subordinated to Manner. However, this proposal must remain tentative until the critical adjective *freely* is more specifically defined. Whether or not the restriction on the formation of the passive voice is related to other restrictions, the fact remains that the construction of the passive is constrained. On the other hand, the main function of the passive, that of reversing the noun phrases, can be accomplished, with some verbs, in the active (Long, 1961, pp. 114–115). Evidence Pairs 14–15 and 16–17.

14. The chairman began the meeting promptly.
15. The meeting began promptly.
16. They ring the bells at six.
17. The bells ring at six.

Finally, psychological studies have shown that the passive is more difficult to understand and to remember (DeVito, 1969; Huttenlocher, Eisenberg, & Strauss, 1968; Slobin, 1966; and the experiments reviewed in the next section). There is also some evidence that changes in the memory of sentences are more often from passive to active than vice versa (Mehler, 1963; Prentice, 1966; see also Sachs, 1967, and Slobin, 1968, for some qualifications on this generalization). Developmentally, too, the active precedes the passive (Gaer, 1969; Hayhurst, 1967; Turner & Rommetveit, 1967a, 1967b). On the basis of the linguistic considerations and the psycholinguistic evidence we can clearly assign marked status to the passive and unmarked status to the active. This conclusion is in accord with the general practice of grammarians to consider the active as basic. Long (1961, Ch. 5) has codified the unmarked status of the active by referring to it as the *common voice.*

Having established that the passive voice is marked and the active unmarked, it seems appropriate to ask what semantic functions, if any, are carried by this difference. This question will be considered in the next two sections in the context of a general discussion of the semantic difference between the two voices.

Active-Passive Semantic Differences: On Contextual Contiguity and Emphasis

Do the active and passive completely overlap in meaning or are there any differences? Ziff (1966) asserted on the basis of examples that the two voices differ in meaning but offered no explicit definition of the difference. Chomsky (1965, Footnote 3 on p. 223; see also, Katz & Martin, 1967) also sensed that the passive is more than a stylistic variant of the active, but saw no way to identify the difference within transformational theory as then formulated. Traditional attempts to specify the functions of the passive voice have pointed to one or more of the following considerations (Jespersen, 1964, pp. 120–121): The passive voice is used when the logical subject is unknown, cannot be easily stated, or is self-evident (e.g., He was killed in the war. I was tempted to start dancing.). The agentless passive is very common. In fact, Jespersen estimated that over 70% of the passives in the works of English writers belonged to this category. Svartvik (1966, p. 141) has reached a similar conclusion. When the agent is provided, the passive is used either because the logical object attracts greater interest than the logical subject and hence is given greater emphasis and prominence by making it grammatical subject, or for stylistic purposes to "facilitate the connexion of one sentence with another: He rose to speak, and was listened to with enthusiasm by the great crowd present [Jespersen, 1964, p. 121]." The emphatic and stylistic functions of the agentful passive often coincide as they both tend to reflect the thematic interests of the speaker. The theme of the discourse receives expression by being placed in grammatical subject position, thus, giving it emphasis and providing discourse unity.

Linguistic treatments of voice normally concentrate on the functions of the passive voice, reflecting the position that no reason needs to be given for the use of the active voice beyond the need to convey information. We will now turn to a review of the psychological literature on semantic differences between the active and passive voices. The psychological studies have generally discussed the differences between the two voices without much attention to the asymmetric character of the relation between them. But a reexamination of the results reveals, as will be seen below, differences to be expected from the marked–unmarked relation between the passive voice and the active voice.

Psychological investigators, without much reference to the linguistic thinking in this domain, have also concentrated on the same two functions in their exploration of the differences between the voices: the contextual contiguity function and the emphatic function. Tannenbaum and Williams (1968a) found that the response latency in forming a sentence to describe a simple line drawing (e.g., a train hitting a car) was longer when the subjects had to describe the drawing in the passive voice than when they had to describe it in the active voice. The differences between the two voices widened when the description task was preceded by a preamble about the logical subject of the picture (in above example: a discourse about trains). On the other hand, the difference shrank when the preamble was about the logical object (about cars). These findings suggest that the response readiness of one voice or another can be primed through the manipulation of the topic of discourse. In the absence of such intervention, however, the active voice seems to be in a steady ready-state, as would be expected from its unmarked status.

In another study, Tannenbaum and Williams (1968b) found that the logical subject was a better cue for the recall of the verb of an active sentence previously presented, whereas the logical object was a better cue for the recall of the verb of a passive sentence. The difference between the two types of cues was, however, more than twice as large for the passive as for the active. In view of the marked status of the passive subject and object, it seems reasonable that the difference between them should be weightier than the difference between their unmarked counterparts in the active. Parallel results were obtained by Turner and Rommetveit (1968) in a study with children using picture cues, rather than word cues. They found that pictures of logical objects were more effective prompts for the recall of passive sentences and pictures of logical subjects better prompts for the recall of active sentences. In this study, too, the difference between the two types of cues was proportionately larger for the passive voice than for the active voice. These studies extend the previous Tannenbaum and Williams findings in showing that individuals are inclined to begin a sentence with what is contextually prominent (here the cueing noun phrase or picture) and that the passive is more sensitive to contextual variation than the active. Using an entirely different method, Prentice (1966) obtained results supporting both conclusions. She found that for both actives and passives it was easier to learn to associate a sentence to an extrasentential noun when the first noun of the sentence (the grammatical subject) was a high associate of that noun than when the last noun of the sentence (the grammatical object) was a high associate. But the difference in ease of learning between the two conditions was much larger for the passive than for the active. The tendency for contextual continuity seems also to be reflected in Wright's (1969) finding that the recall of a sentence 5 seconds after presentation was more accurate when the question about it was in the same voice as the initial sentence than when the sentences and questions were mismatched. The asymmetry between the active and the passive is also reflected in this experiment. The detrimental effect of a passive question about an active sentence was considerably larger than the detrimental effect of an active question about a passive sentence.

The experiments reviewed so far show a tendency for congruence between the grammatical subject and the preceding context, thus suggesting that the grammatical subject is selected to reflect what is contextually prominent. The studies to be discussed below attempt to assess directly the emphatic functions of the grammatical subject. These experiments also indicate that the passive is the one that is particularly adaptive to the context; the active is normally available for use unless contextual weight suggests the choice of the passive. Two studies by Johnson-Laird (1968a, 1968b) were guided by the notion that the function of the passive is to make the logical object more prominent by placing it in the position of the grammatical subject. His sentences had the form *Blue follows* (or precedes) *red* and *Red is followed* (or preceded) *by blue*. The subjects' task was basically to find or express correspondences between these sentences and colored strips. The findings indicate that the grammatical subject (in above example: *blue* for first sentence and *red* for second) was assigned a larger area in the strip than the grammatical object (*red* and *blue*, respectively, for first and second sentence). Also, the difference between size of subject and object was larger for the passive than for the active, reflecting again the marked status of the passive. These findings support the idea that the function of the passive is to emphasize the logical object by making it the grammatical subject. However, the generalizations that can be reached from this study are rather limited because only two nouns and verbs were used (and atypical ones at that) and because of the suggestive nature of the task.

Clark (1965), working within the information theoretic framework, has pointed to the higher interest level of the grammatical subject (the logical object of the passive) than of the grammatical object. In particular, he found:

> The beginning of both grammatical forms tends to be more animate and to have less uncertainty than the end. . . . In ordinary terms, people put what they want to talk about (usually animate things of little uncertainty) in the beginning of the sentence [Clark, 1965, p. 369].

Johnson (1967) seems to have provided additional support for Clark's conclusion.

The psychological studies have thus validated the two functions of the agentful passive, intuitively postulated by linguists: (*a*) to provide discourse continuity and (*b*) to impart prominence to the logical object by placing it in grammatical subject position. These intuitions and findings have to be reconciled with the suggestion made by Mihailović (1963) that the logical subject, not object, is really the emphasized one in the passive because it receives heavy stress: "When the actor is shifted to final position, primary stress falls on it, and makes it very prominent [p. 81]." This suggests that there may be different senses in which the logical object and the logical subject are prominent in the passive voice. The prominence of the logical object has been identified via the reviewed studies in terms of contextual relations. The logical object is shifted in the passive to the position of grammatical subject because it represents the dominant interest of the discourse or because it provides for discourse continuity. In what sense is the logical subject prominent in the passive? An answer to this question will be attempted in the next section through an explication of the concept of *presupposition*.

Active–Passive Semantic Differences: On Presupposition

The notion of *presupposition* has some background in philosophy (see Garner, 1971) and in linguistics (see Halliday, 1967) but has recently been elevated to a position of visibility and importance through the work of generative linguists. Presupposition occupied a central position in a recent linguistic conference (Fillmore & Langendoen, 1971) and it is involved in the controversy whether aspects of meaning not accommodated by the deep structure are a property of the surface structure (Chomsky, 1970) or whether all aspects of meaning are concentrated in a semantic base, bypassing the deep structure altogether (Lakoff, 1971; see also Chafe, 1971). Our application of presupposition to the explication of semantic differences between the two voices can, however, be pursued in a neutral manner. In what follows we will first review the essential aspects of the concept and then turn to its application to the semantic differentiation of the two voices.

The notion of presupposition entails a distinction between two aspects of sentential information: the explicit, new, or illocutionary and the implicit, given, or presupposed. In philosophy, the interest centered on the presupposed existence of definite names involved in sentences. Thus, in Sentence 18 the present existence of a king of France is presupposed as given and his baldness asserted as new information.

18. The present king of France is bald.

Philosophers investigated the effects of failure of presupposition on the status of sentences implying them (see Garner, 1971; Hutchinson, 1971). This issue need not concern us here and we can turn directly to the work of linguists on presupposition. Fillmore (1971a, 1971b) has presented insightful analyses of the semantic composition of lexical items in terms of presupposition. Consider, for instance, Sentence 19.

19. Please shut the door.

Because of the verb *shut,* the sentence implies the presupposition that the door is open. The same presupposition is implied by Sentence 20, which is negative.

20. Please don't shut the door.

Or compare the verbs *accuse* and *criticize* in Sentences 21 and 22, respectively.

21. John accused Harry of writing the letter.
22. John criticized Harry for writing the letter.

Both sentences convey the same information, that John thought Harry did something reprehensible by writing the letter. But whereas Sentence 21 presupposes that writing the letter was blameworthy and asserts the association (in John's statement) of Harry with this act, Sentence 22 presupposes that Harry wrote the letter and asserts the negative evaluation of this act. Thus, what is presupposed as given in Sentence 21 is asserted as new in Sentence 22 and vice versa.

Other lexical items, apart from verbs, also carry presuppositional information (see Keenan, 1971), but our main interest here is in the analysis of presuppositions carried by sentential structure. Chomsky (1970) has based his presuppositional analysis of syntactic structures mainly on the examination of questions. He compares, for instance, Sentences 23 and 24.

23. Did John give the book to Bill?
24. Did John give Bill the book?

Under normal intonation, the last word in each of these sentences receives major phonetic stress. Chomsky proposes that the phrase containing the intonation center be considered the assertional focus of the sentence and the presupposition of the sentence be defined as the proposition resulting from the replacement of the focus by a variable. Thus, the presupposition of Sentence 23 would be that John gave the book to someone and that of Sentence 24 would be that John gave Bill something. The postulated difference in the presuppositional structure of the two sentences is reflected in the difference in status of Responses 25 and 26.

25. No, to someone else.
26. No, something else.

Response 25 appears more natural for Sentence 23 and Response 26 more natural for Sentence 24 because in this pairing the responses share the presuppositions of the questions. But 26 as a response to Sentence 23 and 25 as a response to Sentence 24 appear less natural because they deny the presuppositions of their respective questions. Contrastive intonation can change the focus and presupposition of a sentence. Thus, 26 is a natural response to Sentence 23 when *book* is heavily stressed in Sentence 23. However, the same principles hold for contrastive intonation as for normal intonation and our discussion will evolve around the latter case. The essence of Chomsky's proposal is that the phrase containing the most heavily stressed word is the focus or core of the new information asserted by the sentence; the remaining information in the sentential frame is considered to be presupposed as given.

Halliday (1967–1968), although working from a different perspective, has discussed the notion of presupposition in much the same terms as Chomsky (1970), but with an important addition. He points out that every syntactic structure contains new information but the presupposition (the "given" information, in his terminology) is optional for unmarked structures. That is, in unmarked cases, the information focus may extend over the entire sentence.

With this background on presupposition we can turn to an examination of presuppositional differences between the active voice and the passive voice. Consider a passive sentence as in 27 below.

27. The principal had been angered by the parent.

Primary stress is assigned to *parent* making the phrase *the parent* the focus of the assertion. The presupposition, then, is that the principal had been angered by someone. In other words, this sentence presupposes as known or given that the principal had been angered, and it provides the new information that it was the parent who was responsible for this. This distribution of focus–presupposition as inferred from the stress pattern is supported by a consideration of the choices the speaker could have made, but did not make, in expressing his message. If he wanted to impart just the information about the anger of the principal he could have used the more common agentless passive, saying: The principal had been angered. By adding the by-agent, the speaker has focused attention on it, making it the core of the assertion.

What about the active voice? On the basis of stress consideration alone, the active, like the passive, divides into an assertional and presuppositional component. Thus, consider Sentence 28, the active counterpart of Sentence 27, as an example.

28. The parent had angered the principal.

Primary stress makes *the principal* the focus of the assertion, and its replacement by a variable leaves the presupposition that the parent had angered someone. By this analysis, the noun phrase that enters into the assertion in the passive is part of the presupposition in the active, and vice versa. However, we saw that the semantic differentiation between the two voices with respect to the variables of stylistic continuity and emphasis was mainly due to the effects of these variables in the passive, with the active more or less serving as the neutral base, in accordance with its unmarked status. The markedness distinction would suggest that the passive should also carry the main presuppositional load, with the active being more or less presuppositionally neutral. In fact, this is exactly what Halliday has proposed. He has noted specifically that the presuppositional inference is optional in the active voice (Halliday, 1967, p. 215). He made the point that sentences like 29 below can be analyzed either as presupposing that John's seen something and asserting that it was the play or as lacking any presupposition, with the entire sentence functioning as an assertion.

29. John's seen the play.

The active thus offers the speaker a neutral form for conveying information, one he can use when he does not want to emphasize or to presuppose anything in particular.

There has been virtually no psychological experimentation on presupposition, but one of its psychological consequences is fairly obvious. Chomsky's (1970) discussion reviewed above suggests that it is "unnatural" to deny presuppositions. This point underlies the interviewing effectiveness of questions like 30 below.

30. When did you stop beating your wife?

The verb *stop* asserts the cessation of an activity; it presupposes its prior existence. Or take the noun *bachelor*. Its assertional or foreground meaning refers to a person who has never been married, its presuppositional or background meaning concerns an adult male. Fillmore (1971a) has pointed out that the predication in Sentence 31 involves the assertional meaning of *bachelor*, not its presuppositional content.

31. That person is not a bachelor.

There is also experimental evidence that individuals are less likely to notice events inconsistent with a presupposition than they are to notice events inconsistent with an assertion. Hornby (1971) presented subjects with cleft and pseudocleft sentences as in 32–35.

32. It is the girl that is kicking the pumpkin.
33. The one that is kicking the pumpkin is the girl.
34. It is the pumpkin the girl is kicking.
35. What the girl is kicking is the pumpkin.

Both Sentences 32 and 33 presuppose that someone is kicking the pumpkin and assert that that someone is the girl, whereas Sentences 34 and 35 presuppose that the girl is kicking something and assert that that something is the pumpkin. Following the presentation of each sentence, Hornby flashed briefly, in a tachistoscope, a picture misrepresenting either the presupposition or the assertion of the sentence and requested the subjects to indicate whether the picture was a true or false depiction of the sentence. He found that subjects gave more erroneous "true" responses when the picture misrepresented the presupposition than when it misrepresented the assertion. Thus, for instance, a picture of a girl kicking a football elicited more true responses for Sentences 32 and 33 than for Sentences 34 and 35.

The implication of these observations and findings is that information presented on the presuppositional level is more readily accepted and less likely to be questioned than assertional information. In making this generalization as well as any other ones about presupposition, it must be kept in mind that not all presuppositions are alike. In particular, it appears that there is a scale of strength of presuppositional implication. The presuppositions of some sentential structures appear to be less definite than those of others (see Chomsky, 1970, Footnote 27; Geis & Zwicky, 1971; Karttunen, 1971). Thus, for instance, the presupposition that somebody is kicking the pumpkin appears stronger in Sentence 32 above than in the corresponding passive sentence, 36.

36. The pumpkin is kicked by the girl.

Perhaps the presuppositional message of the passive is weaker because it could have been chosen by the speaker to serve the functions of stylistic continuity and emphasis rather than the presuppositional function.

REFERENCES

ALLEN, W. S. *Living English structure.* London: Longmans, 1959. Cited by J. Svartvik, *On voice in the English verb.* The Hague: Mouton, 1966. P. 2.

ANISFELD, M. Subjective approximation of relative letter incidence in pleasant and unpleasant English words. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7**, 33–40.

CHAFE, W. L. Directionality and paraphrase. *Language,* 1971, **47**, 1–26.

CHOMSKY, N. *Syntactic structures.* The Hague: Mouton, 1957.

CHOMSKY, N. *Aspects of the theory of syntax.* Cambridge, Mass.: M.I.T. Press, 1965.

CHOMSKY, N. Deep structure, surface structure, and semantic interpretation. In R. Jakobson & S. Kawamoto (Eds.), *Studies in general and oriental linguistics.* Tokyo: T. E. C. Co., 1970.

CHOMSKY, N., & HALLE, M. *The sound pattern of English.* New York: Harper & Row, 1968.

CLARK, H. H. Some structural properties of simple active and passive sentences. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4**, 365–370.

CLARK, H. H. Linguistic processes in deductive reasoning. *Psychological Review,* 1969, **76**, 387–404.

CLARK, H. H., & CARD, S. R. Role of semantics in remembering comparative sentences. *Journal of Experimental Psychology,* 1969, **82**, 545–553.

DEVITO, J. A. Some psycholinguistic aspects of active and passive sentences. *Quarterly Journal of Speech,* 1969, **55**, 401–406.

FILLMORE, C. J. Types of lexical information. In F. Kiefer (Ed.), *Studies in syntax and semantics.* New York: Humanities Press, 1971. (a)

FILLMORE, C. J. Verbs of judging: An exercise in semantic description. In C. J. Fillmore and D. T. Langendoen (Eds.), *Studies in linguistic semantics.* New York: Holt, Rinehart & Winston, 1971. (b)

FILLMORE, C. J., & LANGENDOEN, D. T. *Studies in linguistic semantics.* New York: Holt, Rinehart & Winston, 1971.

GAER, E. P. Children's understanding and production of sentences. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 289–294.

Garner, R. "Presupposition" in philosophy and linguistics. In C. J. Fillmore & D. T. Langendoen (Eds.), *Studies in linguistic semantics*. New York: Holt, Rinehart & Winston, 1971.

Geis, M. L., & Zwicky, A. M. On invited inferences. *Linguistic Inquiry*, 1971, **2**, 561–566.

Greenberg, J. H. Language universals. In T. A. Sebeok (Ed.), *Current trends in linguistics*. Vol. 3. *Theoretical foundations*. The Hague: Mouton, 1966.

Halliday, M. A. K. Notes on transitivity and theme in English. (Parts 1–3) *Journal of Linguistics*, 1967, **3**, 37–81, 199–244, and 1968, **4**, 179–215.

Hamilton, H. W., & Deese, J. Does linguistic marking have a psychological correlate? *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 127–134.

Hayhurst, H. Some errors of young children in producing passive sentences. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 634–639.

Hornby, P. A. A psychological investigation of presupposition. Paper presented at the meeting of the Linguistic Society of America, Buffalo, New York, July, 1971.

Hutchinson, L. G. Presupposition and belief-inferences. In, *Papers from the Seventh Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 1971. Pp. 134–141.

Huttenlocher, J., Eisenberg, K., & Strauss, S. Comprehension: Relation between perceived actor and logical subject. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 527–530.

Jespersen, O. *Essentials of English grammar*. University, Alabama: University of Alabama Press, 1964. (Originally published: London: Allen & Unwin, 1933.)

Johnson, M. G. Syntactic position and rated meaning. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 240–246.

Johnson, R. C., Thomson, C. W., & Frincke, G. Word values, word frequency, and visual duration thresholds. *Psychological Review*, 1960, **67**, 332–342.

Johnson-Laird, P. N. The choice of the passive voice in a communicative task. *British Journal of Psychology*, 1968, **59**, 7–15. (a)

Johnson-Laird, P. N. The interpretation of the passive voice. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 69–73. (b)

Karttunen, L. Counterfactual conditionals. *Linguistic Inquiry*, 1971, **2**, 566–569.

Katz, J. J., & Martin, E. The synonymy of actives and passives. *Philosophical Review*, 1967, **76**, 476–491.

Katz, J. J., & Postal, P. M. *An integrated theory of linguistic descriptions*. Cambridge, Mass.: M.I.T. Press, 1964.

Keenan, E. L. Two kinds of presupposition in natural language. In C. J. Fillmore & D. T. Langendoen (Eds.), *Studies in linguistic semantics*. New York: Holt, Rinehart & Winston, 1971.

Lakoff, G. On generative semantics. In D. D. Steinberg & L. A. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy*. London: Cambridge University Press, 1971.

Lees, R. B. The grammar of English nominalizations. *International Journal of American Linguistics*, 1960, **26**(3).

Long, R. B. *The sentence and its parts: A grammar of contemporary English*. Chicago: University of Chicago Press, 1961.

Mehler, J. Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 346–351.

Mihailović, L. Some observations on the use of the passive voice. *English Language Teaching*, 1963, **17**, 77–81.

Miller, G. A. *Language and communication*. New York: McGraw-Hill, 1951.

Postal, P. *Aspects of phonological theory*. New York: Harper & Row, 1968.

Prentice, J. L. Response strength of single words as an influence in sentence behavior. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 429–433.

Sachs, J. S. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, 1967, **2**, 437–442.

Slobin, D. I. Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 219–227.

Slobin, D. I. Recall of full and truncated passive sentences in connected discourse. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 876–881.

Svartvik, J. *On voice in the English verb*. The Hague: Mouton, 1966.

Tannenbaum, P. H., & Williams, F. Generation of active and passive sentences as a function of subject or object focus. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 246–250. (a)

Tannenbaum, P. H., & Williams, F. Prompted word replacement in active and passive sentences. *Language and Speech*, 1968, **11**, 220–229. (b)

Turner, E. A., & Rommetveit, R. The acquisition of sentence voice and reversibility. *Child Development*, 1967, **38**, 648–660. (a)

Turner, E. A., & Rommetveit, R. Experimental manipulation of the production of active and passive voice in children. *Language and Speech*, 1967, **10**, 169–180. (b)

Turner, E. A., & Rommetveit, R. Focus of attention in recall of active and passive sentences. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 543–548.

Wright, P. Transformations and the understanding of sentences. *Language and Speech*, 1969, **12**, 156–166.

Ziff, P. The nonsynonymy of active and passive sentences. *Philosophical Review*, 1966, **75**, 226–232.

(Received August 24, 1971)

Psychological Bulletin
1973, Vol. 79, No. 2, 127–141

DEFINING AND TESTING HYPOTHESES IN MULTIDIMENSIONAL CONTINGENCY TABLES

JULIET POPPER SHAFFER[1]

University of Kansas

The log-linear model is described as a framework for analyzing effects in multidimensional contingency tables, that is, tables of frequencies formed by two or more variables of classification. Variables are considered to have nominal categories. A general purpose analysis is proposed for such tables, similar to analysis of variance. Two test procedures are considered: (*a*) maximum likelihood estimation of expected cell frequencies and associated chi-square tests and (*b*) chi-square tests based on logarithms of adjusted cell frequencies. In addition, two multiple-comparison methods, related to the latter approach, are considered as supplementary or alternative procedures.

Since 1935, when the first article dealing with the estimation of interaction in multidimensional contingency tables was published (Bartlett, 1935), a large literature has accumulated concerning the problems of defining and testing complex relationships in such tables. A previous article in this journal (Sutcliffe, 1957) presented the methods developed by Lancaster (1951) based on the partitioning of a total chi-square statistic in a way analogous to the partitioning of a total sum of squares in the analysis of variance. However, the majority of statisticians working in this area have rejected these methods on the grounds that when relationships among variables are defined in an intuitively acceptable manner, the partitioned chi-square values do not correspond to tests of these relationships (see, e.g., Goodman, 1964b).

The purposes of this article are (*a*) to discuss definitions of relationships among variables in contingency tables and (*b*) to describe practical methods of testing hypotheses about the existence of effects of different degrees of complexity. The article is limited to the consideration of variables with nominal categories (i.e., unordered categories), or where order of categories is to be ignored;[2] and it is concerned primarily with significance testing rather than estimation.

THEORETICAL DEVELOPMENT

In a contingency table, observed entities are classified on the basis of a number of variables, each of which consists of a set of mutually exclusive and exhaustive categories. Each combination of possible variable values defines a cell of the table, and the entry in each cell equals the number of entities exhibiting that particular combination of values. In general, with m variables, the contingency table can be represented as an m-dimensional rectangle with cubical cells.

In the literature, related but not identical distinctions have been made between variables with fixed and random marginal distributions, between independent and dependent variables, and among variables with different causal relationships to one another. These distinctions are considered as they become relevant in the course of the presentation. They affect the nature of the analysis considered here in that they bear on the appropriateness of testing various kinds of relationships, but the form of any test, if appropriate, is the same for all cases.

Note that a q-factor analysis of variance design would correspond to a $(q + 1)$-factor contingency table, since reference to the dependent variable is suppressed in analysis of variance, where all effects can be thought of as interactions between the q independent variables and the one dependent variable. Since all variables are represented explicitly in the contingency table, the frequency in each cell cannot be considered equivalent to a score; that is, it is not analogous to a cell mean in the analysis of variance that represents the mean score on the dependent variable. Nonetheless, it is demonstrated that treating transformed cell frequencies as if they were scores produces a formal correspondence between contingency table models and analysis of variance models.[3]

---

[1] Requests for reprints should be sent to Juliet Popper Shaffer, Department of Psychology, University of Kansas, Lawrence, Kansas 66044.

[2] Note that this implies essentially no restriction for variables with two categories.

Notation

The variables in the contingency table are denoted $A, B, C, \ldots$, with a particular category of A denoted A_i, where $i = 1, \ldots, a$; of B denoted B_j, where $j = 1, \ldots, b$; etc. Further notation is defined for a three-factor table; the extension to tables of other dimensionality should be obvious. The probability that an entity appears in the cell characterized by level i of A, level j of B, and level k of C is denoted p_{ijk}, with $\sum_i \sum_j \sum_k p_{ijk} = 1$.[4] Tables formed by summing p_{ijk} over all categories of one or more variables are called marginal tables. The variables summed over are represented by dots; thus, the entires in the AB marginal table are the values of $p_{ij.}$, those in the A marginal table, the values of $p_{i..}$, etc. These entries could be probabilities, fixed proportions, or some combination of the two, depending upon the method of controlling the variables in the experimental situation. For example, if A represented sex, and equal numbers of men and women were chosen for an investigation, the $p_{i..}$, where $i = 1, 2$, would be fixed proportions, each equal to $\frac{1}{2}$. On the other hand, if the total sample were chosen randomly from a population and sex were merely recorded, the $p_{i..}$ would be probabilities. If the sample were controlled on other characteristics but not on sex, the $p_{i..}$ would be averages of probabilities within subpopulations. The observed frequency in the cell characterized by level i of A, level j of B, and level k of C is denoted N_{ijk}, with the total frequency, $\sum_i \sum_j \sum_k N_{ijk}$, denoted by N. Assuming a fixed value of N, the expected frequency in each cell, Np_{ijk}, is denoted E_{ijk}. The dot notation used for probabilities is used similarly for observed and expected frequencies; thus, $E_{ij.} = \sum_k E_{ijk}$, $N_{i..} = \sum_j \sum_k N_{ijk}$, etc. Maximum likelihood estimates of the expected frequencies in each cell are denoted M_{ijk}, with the dot notation used as for N and E. Any term enclosed in parentheses refers to the set of such values in the entire table; for example, $(N_{ij.})$ denotes the complete set of marginal AB frequencies.[5]

Fixed and Random Marginal Tables

All methods presented are appropriate for tables in which (*a*) marginal distributions on all subsets of variables are random and (*b*) the marginal distribution of some subset of from 1 to $m - 1$ variables is fixed. (In the latter case, all marginal distributions of variables contained in the subset are also necessarily fixed.) Test procedures are suitable for testing relationships among variables not included in the subset with a fixed marginal distribution, and for testing relationships among variables of which some are and some are not in that subset. In general, they are not appropriate for testing relationships among variables which are all in the subset with a fixed marginal distribution.

Although many of the procedures are suitable under more general conditions on marginal tables, Condition *a* or *b* (set forth above) should be satisfied in a large majority of empirical studies.

Log-Linear Model

The log-linear model, with properties investigated extensively by Birch (1963), Bishop (1969), and Haberman (1970), provides a satisfactory framework for a general analysis of contingency tables. In this model, the logarithm of a cell probability is expressed as the sum of a number of effects in an expression formally identical to that for the population value of a cell mean in an analysis of variance design. For example, in a three-factor design, the log-linear model can be written in the following form:

---

[3] Logit analysis and its extensions, which deal with an m-dimensional contingency table where one variable is considered the dependent variable and the other $m - 1$ are considered independent variables, are not considered separately here, since these analyses can be shown to be special cases of the general analysis that is presented (Bishop, 1969).

[4] Summations and products will always be over the full range of possible values of the variables. Therefore, for convenience, the range is not indicated in the summation or product notation.

[5] It is assumed implicitly throughout that all maximum likelihood estimates and expected cell frequencies are greater than zero.

$$\log p_{ijk} = v + v_{1i} + v_{2j} + v_{3k} + v_{12ij} + v_{13ik} + v_{23jk} + v_{123ijk}, \quad [1]$$

where $i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, c$; with linear restrictions on the parameters identical to those in analysis of variance; for example, $\sum_i v_{1i} = \sum_j v_{2j} = \sum_k v_{3k} = 0$, etc.[6] Each hypothesis then can be expressed by setting the appropriate terms equal to zero; for example, the hypothesis $A \times B \times C = 0$ is equivalent to the hypothesis $v_{123ijk} = 0$, for $i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, c$.

In order to clarify the nature of the model, it is discussed first as it applies to contingency tables where each factor has two categories.

Dichotomous Variables

Two factors: The 2 × 2 table. In a 2 × 2 table, the effects that can be considered are A, B, and $A \times B$. Taking the latter first, the hypothesis that the interaction effect is equal to zero in the log-linear model is equivalent to the hypothesis of independence of the two factors; that is, that $p_{ij} = p_{i.}p_{.j}$; where $i, j = 1, 2$. The main effect of variable A can be defined either marginally or conditionally. The marginal main effect would be a function of the marginal proportions, and the hypothesis of zero marginal main effect could be stated as follows: $p_{i.} = \frac{1}{2}$, where $i = 1, 2$. This obviously could be meaningful only if $p_{i.}$ were not a fixed proportion, for any i. Alternatively, the main effect of A could be defined as some type of average of its effects within each level of B, which are referred to as partial effects. The hypothesis that a particular effect equals zero in the log-linear model for the complete table implies an average partial definition of the effect. A definition of partial A effects consistent with the model is $\log (p_{1j}/p_{2j})$, although any function of (p_{1j}/p_{2j}) would be sufficient. The average partial effect of A equals zero in the model if and only if $\sum_j \log (p_{1j}/p_{2j}) = 0$.

Three factors: The 2 × 2 × 2 table. Main effects in the 2 × 2 × 2 table are defined in the same way as in the 2 × 2 table. With three factors, each main effect can be defined marginally with respect to the two other factors, partially with respect to both, or marginally with respect to one and partially with respect to the other. The three two-variable interactions can each be of two types: marginal with respect to the third variable, or an average of the two partial interactions, each at one of the values of the third variable. The hypothseis of zero marginal interaction between A and B can be expressed in the following form: $p_{ij} = p_{i..}p_{.j.}$, where $i, j = 1, 2$. In order to define a zero average partial interaction, a measure of two-variable interaction and a specification of the averaging process is needed. The hypothesis of zero three-variable interaction, as in analysis of variance, is taken to be equivalent to the hypothesis that the two partial two-variable interactions are equal, for every pair of the three variables.

To be consistent with the log-linear model, it is sufficient to define the measure of two-variable interaction as a function of the cross-product ratio $p_{11}p_{22}/p_{12}p_{21}$ in a 2 × 2 table, or $p_{11k}p_{22k}/p_{12k}p_{21k}$ for each of the partial interactions in a 2 × 2 × 2 table. Edwards (1963), Altham (1970), and Mosteller (1968) have presented compelling arguments for using a measure of association based on the cross-product ratio. The hypothesis that $A \times B = 0$ in the log-linear model is equivalent to the hypothesis that $\sum_k \log (p_{11k}p_{22k}/p_{12k}p_{21k}) = 0$. The hypothesis that the triple interaction is zero is equivalent to the hypothesis that the two partial interactions are equal or that $(p_{111}p_{221}/p_{121}p_{211}) = (p_{112}p_{222}/p_{122}p_{212})$. This definition of zero $A \times B \times C$ interaction is equivalent to that given by Bartlett (1935), and has been defended from an information-theoretic point of view by Good (1963)[7] and from a Bayesian point of view by Lindley (1964).

More than three dimensions: The general 2^m table. The extension to more than three dimensions is straightforward. In a four-dimensional table, for example, the hypothesis of zero $A \times B \times C \times D$ interaction is the hypothesis that the two partial three-factor interactions are equal for all combinations of three of the four variables. The three-factor interactions are measured as functions of an extended cross-product ratio; thus, the measure of three-variable interaction in a 2 × 2 × 2 table would be a function of $(p_{111}p_{221}p_{122}p_{212})/(p_{121}p_{211}p_{112}p_{222})$, and the hypothesis of zero four-factor interaction in a 2 × 2 × 2 × 2 table could be expressed as follows:

---

[6] The logarithms can be taken to any base for purposes of the discussion in this section.

[7] Kullback and his coworkers, using information-theoretic concepts, arrive at a related but somewhat different general approach (see, e.g., Ku, Varner, & Kullback, 1971; see also Goodman, 1969b, 1970, 1971a, for discussion and criticism).

$$(p_{1111}p_{2211}p_{1221}p_{2121}/p_{1211}p_{2111}p_{1121}p_{2221}) = (p_{1112}p_{2212}p_{1222}p_{2122}/p_{1212}p_{2112}p_{1122}p_{2222}).$$

General m-Dimensional Contingency Table: Extension of Definitions

In analysis of variance factorial designs, an $A \times B \times C$ interaction is defined to be zero when it is zero for every combination of two levels of A, two levels of B, and two levels of C; an $A \times B$ interaction is defined to be zero when it is zero for every combination of two levels of A and two levels of B, etc. The same definitions apply in the contingency table. It is not necessary to examine all possible subtables in each case. A subset of subtables can be chosen such that if the required relationships hold within the subset, they hold for the set of all relevant subtables for each effect. A convenient way of choosing a subset is to consider only combinations of two levels with adjacent numerical values. For the $A \times B \times C$ interaction in the three-factor table, then, it would be sufficient to consider the $(a - 1)(b - 1)(c - 1)$ subtables formed by taking all combinations of levels i and $(i + 1)$ of A, where $i = 1, \ldots, a - 1$; j and $(j + 1)$ of B, $j = 1, \ldots, b - 1$; and k and $(k + 1)$ of C, where $k = 1, \ldots, c - 1$. For example, in a 2 × 2 × 3 table with variables A, B, and C, three 2 × 2 × 2 subtables can be formed: (*a*) A_{12}, B_{12}, C_{12}; (*b*) A_{12}, B_{12}, C_{13}; and (*c*) A_{12}, B_{12}, C_{23}, where the subscripts indicate the levels of the variable that are present in the table. However, it would be sufficient to consider only the Subtables *a* and *c*. The definition of zero $A \times B \times C$ interaction therefore can be expressed as follows:

$$[p_{ijk}p_{(i+1)(j+1)k}/p_{i(j+1)k}p_{(i+1)jk}] = [p_{ij(k+1)}p_{(i+1)(j+1)(k+1)}/p_{i(j+1)(k+1)}p_{(i+1)j(k+1)}]$$

for $i = 1, \ldots, a - 1$; $j = 1, \ldots, b - 1$; $k = 1, \ldots, c - 1$. Similarly, the $A \times B$ interaction is defined as zero when

$$\log [p_{ijk}p_{(i+1)(j+1)k}/p_{i(j+1)k}p_{(i+1)jk}] = 0.$$

Relationship between Average Partial and Marginal Effects

Analysis of a marginal effect, using marginal contingency tables, should be considered only if the variables to be disregarded are randomly sampled; otherwise, the marginal table would not contain frequencies each of which could be regarded as sampled from a single population. Although the analysis of variance analogues of average partial and marginal effects are always equal in a balanced design, the effects are generally not equal in a contingency table, and the relationship between them is rather complex. Therefore, it is of some importance to have criteria for deciding which effects should be examined. This issue is considered in the context of a three-dimensional contingency table, with Factors A, B, and C, where the $A \times B$ interaction is of interest.

An analogy between average partial $A \times B$ interaction and partial correlation $(\rho_{AB.C})$ and marginal $A \times B$ interaction and marginal correlation (ρ_{AB}) is helpful in understanding the issues in this decision. As in the correlational situation, it seems necessary to consider the meanings of the variables, and in particular, their possible causal interrelationships. Where there is no possibility that C could be an effect of A or B, the average partial $A \times B$ interaction would generally be more meaningful. Thus, if C were an independent variable in the experimental sense, a set of experimentally manipulated conditions, then the average partial $A \times B$ interaction would have a clear interpretation. Similarly, if C were a variable, such as sex, race, or social class (for children), and if the interaction were between two behavioral measures, the partial interaction would be indicated, since these background variables could in no way be functions of the behavior.

On the other hand, if C were a variable that could itself be affected by A or B, such as another behavioral measure, holding C constant could distort the estimated relationship between A and B, as in the case of partial correlation. Simpson (1951) discussed some inter-

TABLE 1

LOG-LINEAR EXPRESSIONS OF HYPOTHESES IN A THREE-FACTOR CONTINGENCY TABLE

Hypothesis	Expression in log-linear model	
	Equations	*df*
$A = 0$	$\sum_i \sum_k \log p_{ijk} - \sum_i \sum_k \log p_{(i+1)jk} = 0$	$a - 1$
$B = 0$	$\sum_i \sum_k \log p_{ijk} - \sum_i \sum_k \log p_{i(j+1)k} = 0$	$b - 1$
$C = 0$	$\sum_i \sum_j \log p_{ijk} - \sum_i \sum_j \log p_{ij(k+1)} = 0$	$c - 1$
$AB = 0$	$\sum_k \log p_{ijk} + \sum_k \log p_{(i+1)(j+1)k} - \sum_k \log p_{i(j+1)k} - \sum_k \log p_{(i+1)jk} = 0$	$(a-1)(b-1)$
$AC = 0$	$\sum_j \log p_{ijk} + \sum_j \log p_{(i+1)j(k+1)} - \sum_j \log p_{ij(k+1)} - \sum_j \log p_{(i+1)jk} = 0$	$(a-1)(c-1)$
$BC = 0$	$\sum_i \log p_{ijk} + \sum_i \log p_{i(j+1)(k+1)} - \sum_i \log p_{ij(k+1)} - \sum_i \log p_{i(j+1)k} = 0$	$(b-1)(c-1)$
$ABC = 0$	$\log p_{ijk} + \log p_{i(j+1)(k+1)} + \log p_{(i+1)j(k+1)} + \log p_{(i+1)(j+1)k} - \log p_{ij(k+1)}$ $- \log p_{i(j+1)k} - \log p_{(i+1)jk} - \log p_{(i+1)(j+1)(k+1)} = 0$	$(a-1)(b-1)(c-1)$

Note.—When equations require summing over indices, their ranges are $i = 1, \ldots, a$; $j = 1, \ldots, b$; $k = 1, \ldots, c$. When equations do not require summing over indices, their ranges are $i = 1, \ldots, a - 1$; $j = 1, \ldots, b - 1$; $k = 1, \ldots, c - 1$.

esting examples illustrating the issues just considered. In many empirical situations the choice between the analysis of average partial and marginal effects may not be clear-cut, and often both may be of interest.

The conditions under which average partial and marginal interactions are identical are difficult to specify in general, but relatively easy to specify for two-factor interaction in a three-factor table. In particular, a zero average partial $A \times B$ interaction implies zero marginal $A \times B$ interaction if and only if either (*a*) $A \times B \times C = 0$, and $A \times C$ (average partial) $= 0$; or (*b*) $A \times B \times C = 0$, and $B \times C$ (average partial) $= 0$. For a detailed discussion of the relationship between partial and marginal effects in general contingency tables, see Goodman (1970, 1971b).

Functions of Cell Probabilities in the Log-Linear Model

Just as in the case of population cell means in analysis of variance, hypotheses that specified effects are zero can be expressed in terms of constraints on cell probabilities in the log-linear model. Table 1 gives the contingency table model equations corresponding to each hypothesis which could be tested in a three-factor contingency table. If $\log p_{ijk}$, $\log p_{ij(k+1)}$, etc., are replaced by μ_{ijk}, $\mu_{ij(k+1)}$, etc., where μ_{ijk} represents the population mean of the dependent variable in cell A_i, B_j, C_k in a factorial analysis of variance design, all expressions would be appropriate to the factorial design.

To extend the results to general contingency tables, a simple rule for generating the required equations is stated. First, consider the 2^m subtable consisting of Levels 1 and 2 for each factor. Given any specific effect, the subscripts of p (or μ) referring to variables represented in that effect will be designated relevant to that effect. For example, subscripts i and j are relevant to effect $A \times B$; i and k are relevant to effect $A \times C$; i, j, and k are relevant to $A \times B \times C$; etc. Take the sum of the $\log p_{ijk}$ (or μ_{ijk}) over all the irrelevant subscripts. To decide whether each of these sums should be added or subtracted in the general expression, add the values of the relevant subscripts of each to obtain the relevant sum. If the effect is an interaction among an even number of variables, put a plus sign in front of each term with an even relevant sum and a minus sign in front of each term with an odd relevant sum. If the effect is an interaction among an odd number of variables, put a plus sign in front of each term with an odd relevant sum and a minus sign in front of each term with an even relevant sum. Set the expression thus generated equal to zero. For example, in the equation for $A \times B = 0$ in Table 1, the relevant subscript sums of successive terms are 2, 4, 3, and 3, respectively—the first two even, and the last two odd. The set of equations appropriate to all levels of the factors then can be obtained by replacing the relevant subscripts by i if equal to 1, or $(i + 1)$ if equal to 2; j if equal to 1, or $(j + 1)$ if equal to 2; etc., according to the positions of the subscripts.

The degrees of freedom for the test of each hypothesis using the techniques to be presented are the same as for the test of the equivalent analysis of variance hypothesis.

PROCEDURES

Two large-sample procedures for testing the significance of effects are presented: chi-square tests based on maximum likelihood estimation of expected cell frequencies and chi-square tests based on logarithms of adjusted observed frequencies. The sample sizes required for their appropriate use remain to be determined by future research. Corresponding exact small-sample tests, illustrated using the contingency table in Table 3, are presented in Shaffer (1972).

In order to provide the rationale for these tests, it is convenient to begin by considering the distinction between hierarchical and nonhierarchical hypotheses.

Hierarchical and Nonhierarchical Hypotheses

In analysis of variance designs, each main effect and interaction effect is generally tested individually for significance. Another possibility is to test simultaneously the hypotheses that a particular effect or set of effects and all higher-order interactions involving the same sets of variables are zero. This test is sometimes recommended in a multifactor analysis of variance design where individual higher-order interactions are not of interest, or where one wishes to avoid finding chance significant higher-order effects simply as a result of the large number of tests otherwise performed (see Anderson, 1968). A hierarchical hypothesis is defined as a hypothesis that a particular effect and all higher-order interaction effects in which the same combination of variables appears are zero; for example, in a four-factor table, the hierarchical hypothesis $A \times B = 0$ includes the individual hypotheses $A \times B = 0$, $A \times B \times C = 0$, $A \times B \times D = 0$, and $A \times B \times C \times D = 0$, implying, therefore, that all partial $A \times B$ interactions, as well as their average, are equal to zero or, in other words, that A and B are independent at all levels of C and D. Another hypothesis of possible interest in a four-factor problem is the hypothesis that all interactions of three or more factors equal zero. To test this would require a simultaneous test of the hypotheses $A \times B \times C = 0$, $A \times B \times D = 0$, $A \times C \times D = 0$, $B \times C \times D = 0$, and $A \times B \times C \times D = 0$. These form a hierarchical set; for every effect equated to zero, all higher-order interactions involving the same combination of variables are also equated to zero.

While each chi-square procedure can be used to test both hierarchical and nonhierarchical hypotheses, the method based on expected cell frequencies is more easily adapted to testing hierarchical hypotheses, while the method based on logarithms of observed cell frequencies is relatively simpler in testing nonhierarchical hypotheses. Therefore, each method is presented in the context in which it is most usefully applied.

Hierarchical Hypotheses: Approximate Chi-Square Tests Based on Maximum Likelihood Estimation of Expected Cell Frequencies

The chi-square test of a hierarchical hypothesis (or set of hierarchical hypotheses) presented here utilizes a maximum likelihood estimate of the expected frequency in each cell, given that the hypothesis is true. The hypothesis can then be tested using an ordinary Pearson chi-square test, which, in the context of a three-factor table, for example, involves computation of the following:

$$X_P^2 = \sum_i \sum_j \sum_k \left[\frac{(N_{ijk} - M_{ijk})^2}{M_{ijk}}\right]. \qquad [2]$$

An alternative chi-square test of the same hypothesis, based on the logarithm of the likelihood ratio, requires the computation of the following:

$$X_L^2 = 2 \sum_i \sum_j \sum_k N_{ijk} \log_e\left(\frac{N_{ijk}}{M_{ijk}}\right). \qquad [3]$$

The two chi-square tests are asymptotically equivalent (Mood & Graybill, 1963). There is no clear evidence that either has an advantage over the other for a moderate number of observations. The less familiar likelihood ratio chi-square test has been introduced because tests of conditional hypotheses, described in a subsequent section, are more readily handled by the use of these chi-square tests.

The extension of both tests to tables with more than three factors should be obvious. The number of degrees of freedom for chi-square equals the number of cells minus the number of linearly independent restrictions employed in obtaining the estimated expected frequencies This number can be arrived at conveniently by treating the hypotheses as though they were the equivalent analysis of variance hypotheses and obtaining the degrees of freedom of the analysis of variance sum of squares that would be appropriate for the test of the composite hypothesis. Thus, in an ABC table, the hypothesis $A \times B \times C = 0$ would be tested approximately by a chi-square test with $(a-1)(b-1)(c-1)$ degrees of freedom; the hierarchical hypothesis $A \times B = 0$, which would imply both $A \times B \times C = 0$ and $A \times B = 0$, would be tested by a chi-square test with $(a-1)(b-1)(c-1) + (a-1)(b-1)$ degrees of freedom, etc. Generally, the larger the set of individual hypotheses, the easier the determination of maximum likelihood estimates of expected frequencies. In some but not all cases, explicit expressions for these estimates in terms of marginal frequencies can be obtained.

The general procedure is illustrated first for a three-factor table. Table 2 presents every possible combination of hierarchical hypotheses in a three-factor table (except those that can be obtained by symmetry), the appropriate formulas for estimating expected frequencies, and the degrees of freedom for the chi-square test. In a three-factor table, the hypothesis $A \times B \times C = 0$ is the only one that does not result in an explicit expression for estimates of expected cell frequencies. These estimates can be obtained in a straightforward manner only by solving a set of $(a-1)(b-1)(c-1)$ simultaneous fourth-degree equations (Goodman, 1964b).

As the number of variables increases, so does the number of hierarchical hypotheses that do not yield explicit expressions for maximum likelihood estimates of cell frequencies. Goodman referred to these as nonelementary, as opposed to elementary, hypotheses (Goodman, 1970). As Table 2 indicates, only one of the eight types of hypotheses in a three-factor table is nonelementary, while in a four-factor table, 10 of the 27 types are nonelementary (see Goodman, 1970, p. 239, for a full list of hierarchical hypotheses in a four-factor table). As the number of factors increases, the complexity of straightforward estimation also increases. The adaptation, during the 1960s, of a relatively simple iterative method to obtain these expected frequencies marked a real breakthrough in providing practical methods for analyzing multidimensional tables. The method is based on the fact that when maximum likelihood estimates of expected frequencies under a given hierarchical hypothesis are substituted

TABLE 2

HIERARCHICAL HYPOTHESES IN A THREE-FACTOR TABLE

Hierarchical hypotheses	Effects equal to zero	Expression for M_{ijk}	df
$A = 0$	A, $A \times B$, $A \times C$, $A \times B \times C$	$N_{.jk}/a$	$(a-1)bc$
$A = 0$, $B = 0$	A, B, $A \times B$, $A \times C$, $B \times C$, $A \times B \times C$	$N_{..k}/ab$	$(ab-1)c$
$A = 0$, $B = 0$, $C = 0$	A, B, C, $A \times B$, $A \times C$, $B \times C$, $A \times B \times C$	N/abc	$abc - 1$
$A = 0$, $B \times C = 0$	A, $A \times B$, $A \times C$, $A \times B \times C$, $B \times C$	$N_{.j}N_{..k}/Na$	$abc - b - c + 1$
$A \times B = 0$	$A \times B$, $A \times B \times C$	$N_{i.k}N_{.jk}/N_{..k}$	$(a-1)(b-1)c$
$A \times B = 0$, $A \times C = 0$	$A \times B$, $A \times B \times C$, $A \times C$	$N_{i..}N_{.jk}/N_{...}$	$(a-1)(bc-1)$
$A \times B = 0$, $A \times C = 0$, $B \times C = 0$	$A \times B$, $A \times C$, $B \times C$, $A \times B \times C$	$N_{i..}N_{.j.}N_{..k}/N^2$	$abc - a - b - c + 2$
$A \times B \times C = 0$	$A \times B \times C$	No explicit expression	$(a-1)(b-1)(c-1)$

TABLE 3

CONTINGENCY TABLE FOR TEST OF HYPOTHESIS $A \times B \times C = 0$

Contingency table

Category	$C = 1$		$C = 2$		$C = 3$	
	$B = 1$	$B = 2$	$B = 1$	$B = 2$	$B = 1$	$B = 2$
$A = 1$	0	4	3	7	7	3
$A = 2$	5	1	2	1	1	1

Marginal tables

	$B = 1$	$B = 2$	$C = 1$	$C = 2$	$C = 3$
$A = 1$	10	14	4	10	10
$A = 2$	8	3	6	3	2
$B = 1$			5	5	8
$B = 2$			5	8	4

for observed frequencies, marginal tables based on these estimated expected frequencies are the same as those based on the observed frequencies for all combinations of variables corresponding to effects that are not assumed to be zero according to the hypothesis.

The approach is illustrated first for the test of the hypothesis $A \times B \times C = 0$ in the contingency table given in Table 3. Its extension to the general case then is indicated. Fortunately, the method is particularly simple when applied to elementary hypotheses, so that it is practical to apply it to find estimated frequencies under any hierarchical hypothesis without the necessity of first determining whether or not it is elementary.

Since the hypothesis implies only that $A \times B \times C$ equals zero, the marginal frequencies AB, AC, and BC are fixed; that is, the expected frequencies, substituted for the observed frequencies, will yield the same AB, AC, and BC marginal tables as the observed frequencies. (The A, B, and C marginal tables are fixed as a consequence of having fixed AB, AC, and BC marginal frequencies.) Write out an initial contingency table with 1.0 in every cell. A series of iterations is used to change these preliminary values into expected frequencies (see Table 4 for an illustration of the procedure). Each cycle, in this example, consists of three steps, one corresponding to each set of fixed marginal frequencies:

1. Adjust AB frequencies: Multiply the number in each cell with level i of A and level j of B by the fixed A_iB_j marginal frequency divided by the sum of the current numbers in all these cells. If this is carried out for each combination of i and j values, the new cell numbers will have the correct AB marginal frequencies.
2. Adjust AC frequencies: Using the new cell numbers obtained in Step 1, multiply the number in every cell with level i of A and level k of C by the fixed A_iC_k marginal frequency, divided by the sum of the current numbers in all these cells. Do this for each combination of i and k values; this step yields new cell numbers with the correct AC marginal frequencies.
3. Adjust BC frequencies: Using the cell numbers obtained in Step 2, multiply the num-

TABLE 4

OBTAINING (M_{ijk}) GIVEN $A \times B \times C = 0$ FOR THE CONTINGENCY TABLE IN TABLE 3

Category of A	$C = 1$		$C = 2$		$C = 3$	
	$B = 1$	$B = 2$	$B = 1$	$B = 2$	$B = 1$	$B = 2$
Initial table						
$A = 1$	1	1	1	1	1	1
$A = 2$	1	1	1	1	1	1
Cycle 1, Step 1: Adjust AB frequencies						
$A = 1$	3.333	4.667	3.333	4.667	3.333	4.667
$A = 2$	2.667	1.000	2.667	1.000	2.667	1.000
Cycle 1, Step 2: Adjust AC frequencies						
$A = 1$	1.667	2.333	4.167	5.833	4.167	5.833
$A = 2$	4.364	1.636	2.182	.818	1.455	.546
Cycle 1, Step 3: Adjust BC frequencies						
$A = 1$	1.382	2.939	3.282	7.016	5.930	3.658
$A = 2$	3.618	2.061	1.718	.984	2.070	.342
Cell entries after 7 cycles: (M_{ijk})[a]						
$A = 1$	.975	3.027	2.851	7.149	6.178	3.822
$A = 2$	4.025	1.973	2.149	.851	1.822	.178

[a] $\chi^2 = 6.50$, $df = 2$, $p = .04$.

ber in every cell with level j of B and k of C by the fixed B_jC_k marginal frequency, divided by the sum of the current numbers in all these cells. If this is done for each combination of j and k values, the new cell entries will have the correct BC marginal frequencies.

Steps 1–3 constitute a single cycle in the iterative method. In general, each step results in cell entries satisfying the marginal restrictions adjusted for in that step, but no longer satisfying marginal restrictions adjusted for in previous steps. If the original cell entries yield proportions satisfying the hypothesis to be tested (this can be assured for hypotheses specifying zero effects by taking all initial cell entries equal), then, at every step in each cycle the obtained cell entries yield proportions satisfying the hypothesis.

If the same cycle of three steps is repeated successively on the cell entries obtained at the end of each preceding cycle, the cell entries approach the estimated expected frequencies, given the hypothesis. At the end of each cycle, the marginal frequencies obtained from the cell entries are compared with the required marginal frequencies. When the two sets match within a desired degree of accuracy, the process is terminated, and the cell entries at that point are taken as the estimated expected frequencies. Apparently, a small number of cycles are sufficient to obtain a highly accurate approximation: 5–7 cycles have been found sufficient in most cases (Ku, Varner, & Kullback, 1971). In the example illustrated in Table 4, seven cycles were required to obtain agreement within .1% between each obtained and required marginal frequency.

In the case of an elementary hypothesis, agreement between obtained and required marginal frequencies is perfect (within limits of computational accuracy) at the end of the first cycle.[8] Thus, the failure to recognize an elementary hypothesis and to use the explicit formula does not lead to any substantially greater computational effort.

The set of marginal frequencies to be fixed, that is, to be adjusted for, in the test of any hierarchical hypothesis or set of hypotheses in a general contingency table, can be determined by following a procedure that is illustrated here for a particular example. Suppose, in a four-factor table, a test of the pair of hierarchical hypotheses $A \times B = 0$ and $B \times C \times D = 0$ is desired. The full set of hypotheses is, then, $A \times B = 0$, $A \times B \times C = 0$, $A \times B \times D = 0$, $A \times B \times C \times D = 0$, $B \times C \times D = 0$. The marginal tables to be fixed can be obtained as follows: the full set of effects is written out and all those assumed to be zero are eliminated (indicated by line drawn through). All marginal totals that are fixed automatically as a consequence of fixing other marginal tables are enclosed in parentheses. (For example, if the ACD marginal totals are fixed, the AC, AD, CD, A, C, and D marginal frequencies are fixed automatically as a result.) The remaining marginal frequencies (circle) constitute the set that must be adjusted for in determining the expected frequencies. For this example, the list would look as follows: (A), (B), (C), (D), ~~AB;~~ (AC), (AD), $\textcircled{BC,}$ $\textcircled{BD,}$ (CD), ~~ABC;~~ ~~ABD;~~ $\textcircled{ACD,}$ ~~BCD;~~ ~~$ABCD$.~~

For another example, to test the hierarchical hypothesis $A \times B = 0$ from the data in Table 3, only Steps 2 and 3 as enumerated above would be carried out, since the AB marginal frequencies would not be fixed in testing this hypothesis. Since $A \times B = 0$ is an elementary hypothesis, only one cycle would be required; the entries obtained at the end of the cycle would be the same as those which would be obtained using the explicit formula in Table 2.

The general iterative method illustrated can be applied for testing any hierarchical hypothesis. Haberman (1970) and Goodman (1971b) give methods for simplifying the iterative procedure in some cases by taking advantage of particular characteristics of the hypotheses being tested.

The ordinary chi-square test of the hypothesis $A \times B \times C = 0$ in the illustrative contingency table, based on the estimated expected frequencies arrived at in Table 4, yields a p value of about .04. As an approximation to the exact probability of .07, the chi-square value is an underestimate, but the agreement is fairly good, considering the extremely small frequencies. In the absence of any empirical investigations of the chi-square approximation in this context, the general recommendations of Cochran (1952, 1954) on sample size could be taken as a rough guide for determining when the chi-square test would be appropriate.

---

[8] This has been proved to hold for up to seven variables. Beyond that, no case is known where more than two cycles are necessary (see Haberman, 1970).

Relationships among Hierarchical Hypotheses and between Full and Marginal Tables: Conditional Chi-Square Tests

In a factorial three-factor analysis of variance design, the sum of squares for testing the hierarchical hypothesis $A \times B = 0$ (i.e., the individual hypotheses $A \times B = 0$ and $A \times B \times C = 0$) with $(a - 1)(b - 1)(c - 1) + (a - 1)(b - 1)$ degrees of freedom, minus the sum of squares for testing the hypothesis $A \times B \times C = 0$, with $(a - 1)(b - 1)(c - 1)$ degrees of freedom, equals the sum of squares for testing the nonhierarchical hypothesis $A \times B = 0$, with $(a - 1)(b - 1)$ degrees of freedom. An equivalent relationship does not hold in contingency table analysis because the individual effects are not estimated orthogonally. However, Goodman (1969b) has shown that if likelihood ratio chi-square statistics are computed to test the hierarchical hypothesis $A \times B = 0$ and the hypothesis $A \times B \times C = 0$, the difference between them is distributed as chi-square with $(a - 1)(b - 1)$ degrees of freedom if the conditional hypothesis $A \times B = 0$ *given that* $A \times B \times C = 0$ is true. In other words, the difference between the two likelihood ratio chi-square statistics is an appropriate test statistic for testing the nonhierarchical hypothesis $A \times B = 0$ *assuming that* $A \times B \times C = 0$, but is not appropriate for testing that hypothesis if this assumption is not made. Goodman has extended this type of analysis to provide a complete partition of the likelihood ratio chi-square statistic appropriate for testing the hierarchical hypotheses that all factors are independent (i.e., that all interactions are zero) into a set of likelihood ratio chi-square statistics, each of which is appropriate for testing a conditional hypothesis of this sort. Of special interest is the fact that these types of hypotheses provide a bridge between the analysis of full contingency tables and marginal tables. In an earlier section, for example, it was pointed out that the average partial interaction of two factors in a three-factor table is equal to their interaction in the marginal two-factor table if the triple interaction is zero and at least one of the other average partial two-factor interactions is zero. Thus, a test of the hypothesis that the two-factor marginal interaction $A \times B = 0$, carried out by a test on the marginal $A \times B$ table, is equivalent to a test of the hypothesis that the average partial $A \times B = 0$, *given that* $A \times B \times C = 0$ and either $A \times C$ or $B \times C$ equals zero, carried out on the full contingency table. For extensions of this reasoning to higher-order contingency tables, see Goodman (1970, 1971a), who makes use of these relationships to build complex models through stepwise procedures somewhat analogous to those used in regression analysis.

One could claim that hierarchical hypotheses or conditional hypotheses of this type are more meaningful than the unconditional nonhierarchical hypotheses. For instance, in a three-factor table, the interpretation of a significant $A \times B$ interaction is certainly more complicated if there is $A \times B \times C$ interaction. Therefore, it could be claimed that one should first test $A \times B \times C$, and if the effect is not significantly different from zero, it should be assumed to be zero, and the test of the conditional hypothesis $A \times B = 0$, assuming $A \times B \times C = 0$, should be carried out. If the $A \times B \times C$ interaction is significant, interpretations should proceed at this level without performing tests of the simpler relationships.

While this procedure may be appropriate under some circumstances, particularly with a small number of factors, it will often be useful to test the nonhierarchical hypotheses, since the simpler relationships are often much greater in magnitude than the more complex higher-order interactions, even if the latter do reach statistical significance. Under these circumstances, the simpler relationships can usually be interpreted without excessive qualification. Furthermore, the assumption that a higher-order interaction is zero when the appropriate test statistic does not reach statistical significance is often in error, and as a result, the test of the conditional hypothesis concerning the simpler effect is biased to an unknown degree. The analysis of individual nonhierarchical hypotheses, as in analysis of variance, is therefore recommended in general as a first overall analysis, except that hierarchical hypotheses dealing with higher-order interaction effects may be tested simultaneously, as discussed previously.

Nonhierarchical Hypotheses: Tests Based on Logarithms of Adjusted Cell Frequencies

In the log-linear model, Equation 1, $\log p_{ijk}$ is expressed as the sum of a number of effects. If the total number of observations in the contingency table is taken as a fixed number, N, note that the model and the hypotheses expressed in terms of the model can be identically expressed in terms of logarithms of expected frequencies E_{ijk}, where $E_{ijk} = Np_{ijk}$. The formulation in terms of expected frequencies is utilized in the discussion to follow, as it is somewhat more convenient in developing tests.

The assumptions made in analysis of variance models of independence, normality, and homogeneity of variance of observed cell means as estimates of population cell means cannot reasonably be made with respect to estimates of logarithms of cell probabilities or logarithms of expected cell frequencies in the log-linear model. However, for large values of N, if linear functions of logarithms of expected cell frequencies are estimated by the corresponding functions of logarithms of observed cell frequencies, the distributions of these estimates can be closely approximated by treating the logarithms of observed cell frequencies as though they were values of independent, normally distributed random variables, with the variance of each equal to the reciprocal of the cell frequency.[9] These asymptotic properties can be utilized to construct significance tests of each effect individually (i.e., nonhierarchically), as in an analysis of variance design.[10]

The approach is described first in the context of a 2 × 2 × 2 table. The hypothesis of zero three-factor interaction is expressed in terms of logarithms of cell probabilities in Table 1; it can be expressed in terms of logarithms of expected frequences E_{ijk}, for two levels of each factor, as follows:

$$\log E_{111} + \log E_{122} + \log E_{212} + \log E_{221} - \log E_{112} - \log E_{121} - \log E_{211} - \log E_{222} = 0. \quad [4]$$

Similarly, the expression for zero average partial $A \times B$ interaction is as follows:

$$\sum_k \log E_{11k} + \sum_k \log E_{22k} - \sum_k \log E_{12k} - \sum_k \log E_{21k} = 0. \quad [5]$$

In other words, each hypothesis is equivalent to a statement that a single contrast among the $\log E_{ijk}$s is equal to zero.

Rather than using the logarithms and reciprocals of the sample frequencies themselves to estimate the terms and variance of the contrast, a better estimate is obtained by adding $\frac{1}{2}$ to each observed frequency before computing these functions (Goodman, 1970). This will also allow estimates when there are obtained frequencies of zero, for which logarithms and reciprocals are not defined, although if the frequencies are generally small, the methods of this section do not apply. Using N_{ijk}^* to represent $N_{ijk} + \frac{1}{2}$, the adjusted cell frequencies, the sample estimate of the $A \times B \times C$ interaction[11] is

$$S_{ABC} = \log N_{111}^* + \log N_{122}^* + \log N_{212}^* + \log N_{221}^* - \log N_{112}^* - \log N_{121}^* - \log N_{211}^* - \log N_{222}^*,$$

and the sample estimate of its variance is the sum of the estimated variance of each term, or

$$V_{ABC} = \sum_i \sum_j \sum_k (1/N_{ijk}^*).$$

Then $S_{ABC}/\sqrt{V_{ABC}}$, or the sample estimate divided by its estimated standard deviation, is approximately a unit normal deviate, and the significance of $A \times B \times C$ interaction can be determined by referring the ratio to tables of the unit normal distribution. Equivalently, the ratio S_{ABC}^2/V_{ABC} is distributed approximately as chi-square with one degree of freedom if $A \times B \times C = 0$. Similarly, the average partial $A \times B$ interaction can be tested by computing

---

[9] Asymptotic variance statements apply to natural logarithms. Therefore, these test statistics, as given, require the use of natural logarithms.

[10] Similar asymptotic properties hold for the frequencies themselves, and tests based directly on the frequencies can be developed (Goodman, 1964b). However, these tests are not symmetric with respect to the numbering of categories, and therefore only the tests based on logarithms will be presented here.

[11] As pointed out, it is necessary to assume only that the measure of an effect is any function of the appropriate ratio of products of probabilities. For ease of exposition, these tests will be described as though the logarithmic function had been chosen. The tests are valid, however, regardless of the particular function used to measure the degree of the effect, so long as it is a function of the appropriate ratio.

$$\frac{S_{AB}^2}{V_{AB}} = \frac{(\sum_k \log N_{11k}^* + \sum_k \log N_{22k}^* - \sum_k \log N_{12k}^* - \sum_k \log N_{21k}^*)^2}{\sum_i \sum_j \sum_k (1/N_{ijk}^*)}, \qquad [6]$$

which is distributed approximately as chi-square with one degree of freedom if this effect is zero.

The approach can be extended easily to test any effect that can be specified by a single equation. This would apply to all effects in a 2^m table, and all effects estimated with one degree of freedom in any table. The equation specifying each such effect sets a linear function of logarithms of expected cell frequencies equal to zero. Compute the sample estimate of this function by replacing expected cell frequencies by adjusted cell frequencies N^*. Compute the variance of the function assuming that the sample estimates for the cells are independently distributed with variances equal to the reciprocals of the adjusted frequencies. The square of the sample estimate divided by the variance is distributed approximately as chi-square with one degree of freedom if the hypothesis is true.

The method can be applied without modification to test a hypothesis specifying a fixed value for *any* linear function of the logarithms of the cell probabilities or expected frequencies. For example, in a three-factor table, any hypothesis of the form $\sum_i \sum_j \sum_k \ell_{ijk} \log E_{ijk} = \ell_0$ can be tested by forming the ratio (S_L^2/V_L), where

$$S_L = \sum_i \sum_j \sum_k \ell_{ijk} \log N_{ijk}^* - \ell_0$$

and

$$V_L = \sum_i \sum_j \sum_k \ell^2_{ijk}/N_{ijk}^*.$$

The ratio is distributed as chi-square with one degree of freedom if the hypothesis is true.

Applying the procedure to test the hypothesis $A \times B = 0$ in the $2 \times 2 \times 3$ table previously considered, the resulting values are as follows: $S_{AB} = -4.007$, $V_{AB} = 6.309$, $(S_{AB}^2/V_{AB}) = 2.55$, $p = .11$. The p value differs considerably from the value derived from the exact test, .23. With such small cell frequencies, this approximate method is clearly inappropriate.

The method is very simple when the hypothesis to be tested is equivalent to the statement that a linear function of the logarithms of the expected frequencies has a specified value, but is considerably more complex when the hypothesis implies specific values for a set of such functions. The larger the set, in general, the more laborious the procedure becomes. This problem arises both in testing hierarchical hypotheses and in testing nonhierarchical hypotheses concerning effects that are estimated with more than one degree of freedom. In general, under these conditions, the method based on logarithms of adjusted cell frequencies becomes analogous to a multivariate analysis of variance. There seems to be no point in considering it for testing hierarchical hypotheses, since the method based on maximum likelihood estimation is simpler in that case. At any rate, it would be applied in the same way as in testing a nonhierarchical hypothesis.

The hypothesis to be tested should be specified in terms of a minimum set of functions of logarithms of expected frequencies (or cell probabilities) that are assumed equal to zero, as detailed for the three-factor problem in Table 1. Consider any s such functions, for example, for a three-factor table

$$L_1 = \sum_i \sum_j \sum_k \ell_{1ijk} \log E_{ijk} - \ell_{10},$$

$$L_2 = \sum_i \sum_j \sum_k \ell_{2ijk} \log E_{ijk} - \ell_{20}, \ldots,$$

$$L_s = \sum_i \sum_j \sum_k \ell_{sijk} \log E_{ijk} - \ell_{s0}.$$

The estimated values from the sample would be $S_1, S_2, \ldots, S_s$, where $S_1 = \sum_i \sum_j \sum_k \ell_{1ijk} \times \log N_{ijk}^* - \ell_{10}$, $S_2 = \sum_i \sum_j \sum_k \ell_{2ijk} \log N_{ijk}^* - \ell_{20}$, etc. The estimated variances would be, respectively,

$$V_{11} = \sum_i \sum_j \sum_k \ell_{1ijk}^2/N_{ijk}^*,$$

$$V_{22} = \sum_i \sum_j \sum_k \ell_{2ijk}^2/N_{ijk}^*,$$

etc., and the estimated covariances would be

$$V_{12} = V_{21} = \sum_{i} \sum_{j} \sum_{k} \ell_{1ijk}\ell_{2ijk}/N_{ijk}^*,$$

$$V_{13} = V_{31} = \sum_{i} \sum_{j} \sum_{k} \ell_{1ijk}\ell_{3ijk}/N_{ijk}^*,$$

etc. Let **S** be the column vector of estimated functions, that is, $\mathbf{S} = \text{col}\,(S_1, S_2, \ldots, S_s)$. Let **V** be the (symmetric) covariance matrix, that is,

$$\mathbf{V} = \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1s} \\ V_{21} & V_{22} & \cdots & V_{2s} \\ \vdots & \vdots & \cdots & \vdots \\ V_{s1} & V_{s2} & \cdots & V_{ss} \end{bmatrix} \qquad [7]$$

Then

$$X^2 = \mathbf{S}'\mathbf{V}^{-1}\mathbf{S} \qquad [8]$$

would be distributed approximately as chi-square with s degrees of freedom if the hypothesis were true (Anderson, 1958).

In general, the procedure requires computation and inversion of a covariance matrix of order equal to the number of degrees of freedom. In many cases, it is possible to simplify the test so that inversion only of one or more matrices of considerably lower order is required (for details see Goodman, 1964a). In particular, in testing the hypothesis that the highest-order interaction equals zero in a table where only one of the variables has more than two categories, it is unnecessary to perform any matrix inversion. The simplified procedure was used to test the hypothesis $A \times B \times C = 0$ in the $2 \times 2 \times 3$ table in which that hypothesis was tested using the chi-square test based on estimated expected frequencies (Table 4). The obtained p value, .15, is not in good agreement with either the exact p of .07 or the p of .04 based on the maximum likelihood chi-square test. The example was presented only for illustration; the cell frequencies in this instance are too small for the appropriate use of this test.

Logarithms of Adjusted Cell Frequencies and Multiple-Comparisons Methods

Two main multiple-comparisons methods are presented here. They can be considered either as alternatives to the test procedure using Equation 8, or, in the case of the first method at least, as supplementary to it. The first is an analog of the Scheffé procedure in analysis of variance (Goodman, 1964a; Scheffé, 1959). Consider, for example, the hypothesis $A \times B \times C = 0$ in a three-factor table. This hypothesis can be specified by setting $(a - 1)(b - 1)(c - 1)$ linear functions of the logarithms of the F_{ijk} equal to zero. Since, if the hypothesis is true, all these functions equal zero, it follows that all linear combinations of them equal zero. This includes measures of interaction in all $2 \times 2 \times 2$ subtables of the full table, and averages of these. Let f_{ABC} indicate a single function of this class. Each such function can be expressed as a linear function of the F_{ijk}, and thus S_f and V_f can be computed, as discussed previously. The probability is approximately $(1 - \alpha)$ that, *for all* f, $S_f^2/V_f < \chi_\alpha; (a - 1)(b - 1)(c - 1)^2$, where $\chi_\alpha; (a - 1)(b - 1)(c - 1)^2$ refers to the chi-square value at the alpha significance level, for $(a - 1)(b - 1)(c - 1)$ degrees of freedom. In other words, if any function f can be found for which

$$\frac{S_f^2}{V_f} > \chi_\alpha; (a - 1)(b - 1)(c - 1)^2, \qquad [9]$$

the hypothesis $A \times B \times C = 0$ can be rejected at the alpha significance level. The advantages and disadvantages of the Scheffé procedure have been widely discussed in the context of analysis of variance. An additional advantage in the contingency table context, when considering it as an alternative to the overall chi-square, is its extreme simplicity as compared to the overall test. One can look at the results and pick a function which appears most likely to be significant, and then test it using this procedure. Alternatively, only a certain set of functions may be of interest, in which case each of them may be tested in this manner, keeping the overall significance level less than or equal to alpha.

In the case where only a finite set of functions is of interest, another possible simultaneous test procedure, which should be considered as an alternative to the overall chi-square test, is based on Bonferroni statistics (Goodman, 1964a; Miller, 1966). If, for example, a set of F functions is of interest and if each is tested individually at significance level α/F, the overall level is $\leq \sum_{f} \alpha/F = \alpha$. (Actually, each could be tested at a different level, as long as the sum over all functions were equal to alpha.)

Applied to the previous example, if f were one of F functions to be considered, it would be considered significantly different from zero if

$$\frac{S_f^2}{V_f} > \chi_{\alpha/F;\,1}^2, \qquad [10]$$

where $\chi_{\alpha/F;\,1}^2$ is the chi-square value at the α/F significance level for one degree of freedom. If any of the F functions were found significant by this criterion, the hypothesis $A \times B \times C = 0$ would be rejected, at the alpha significance level.

If a specific set of functions is of interest, then S_f^2/V_f could be computed for each, and compared with either $\chi_{\alpha;\,(a-1)(b-1)(c-1)}^2$ or $\chi_{\alpha/F;\,1}^2$. The relative power of the two procedures, given a specific contingency table, would depend on the number of functions F. It would be legitimate in any specific situation to compute both chi-square criterion values and to choose the smaller one to obtain the more powerful test.

A set of functions which might form a reasonable finite set for general purpose hypothesis testing would be the set of interaction effects in all $2 \times 2 \times 2$ subtables. In that case, F would be equal to

$$\binom{a}{2}\binom{b}{2}\binom{c}{2}$$
$$= a(a-1)b(b-1)c(c-1)/(2 \times 2 \times 2).$$

This would be somewhat analogous to considering differences between all pairs of means in a simple analysis of variance design. For the usual significance levels, .05 and .01, the method based on Bonferroni statistics generally gives a more powerful test than the Scheffé-type alternative with respect to this set of functions (a useful table for comparing the two can be found in Goodman, 1969a, p. 11). This implies that the method based on Bonferroni statistics is more powerful than the overall chi-square test with respect to these functions. Considering the much greater simplicity of the Bonferroni tests, and the fact that the information gained may be more readily interpreted, this method can be considered a serious competitor to the overall chi-square test. Its advantages with respect to the latter are analogous to the advantages of the Tukey studentized range test with respect to the F test, with the additional advantage of much greater ease of computation.

Although the multiple-comparisons methods have been described in the context of the test of $A \times B \times C$ interaction in a three-factor table, the principles can be readily extended to test any hypothesis in a general table.

CONCLUSIONS

The most pressing requirement at this stage is research to indicate what conditions on sample size are necessary in order to get reasonably close approximations using the chi-square techniques. These conditions are likely to be considerably more complex than for two-factor contingency tables. Probability levels obtained by chi-square procedures should be compared with exact probability levels, based on conditional tests, and with unconditional probability levels derived from sampling from populations under various conditions. Modifications to obtain more satisfactory approximations with moderate sample size should be systematically explored; many variations have been suggested in the literature.

Although continual refinement of methods is to be expected, the methods that have been presented are sufficiently practical and broad in scope to constitute a highly useful set of techniques for the detailed analysis of multiple-contingency tables.

REFERENCES

ALTHAM, P. M. E. The measurement of association of rows and columns for an r × s contingency table. *Journal of the Royal Statistical Society* (Series B), 1970, **32**, 63–73.

ANDERSON, N. H. Partial analysis of high-way factorial designs. *Behavior Research Methods and Instrumentation*, 1968, **1**, 2–7.

ANDERSON, T. W. *Introduction to multivariate statistical analysis*. New York: Wiley, 1958.

BARTLETT, M. S. Contingency table interactions. *Journal of the Royal Statistical Society* 1935, **2**, 248–252. (Supplement)

BIRCH, M. W. Maximum likelihood in three-way contingency tables. *Journal of the Royal Statistical Society* (Series B), 1963, **25**, 220–233.

BISHOP, Y. M. N. Full contingency tables, logits, and split contingency tables. *Biometrics*, 1969, **25**, 383–399.

COCHRAN, W. G. The χ^2 test of goodness of fit. *Annals of Mathematical Statistics*, 1952, **23**, 315–345.

COCHRAN, W. G. Some methods for strengthening the common χ^2 tests. *Biometrics*, 1954, **10**, 417–451.

EDWARDS, A. W. F. The measure of association in a 2×2 table. *Journal of the Royal Statistical Society* (Series A), 1963, **126**, 109–114.

GOOD, I. J. Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *Annals of Mathematical Statistics*, 1963, **34**, 911–934.

GOODMAN, L. A. Interactions in multidimensional contingency tables. *Annals of Mathematical Statistics*, 1964, **35**, 632–646. (a)

GOODMAN, L. A. Simple methods for analyzing three-factor interaction in contingency tables. *Journal of the American Statistical Association*, 1964, **59**, 319–352. (b)

GOODMAN, L. A. How to ransack social mobility tables and other kinds of cross-classification tables. *American Journal of Sociology*, 1969, **75**, 1–40. (a)

GOODMAN, L. A. On partitioning χ^2 and detecting partial association in three-way contingency tables. *Journal of the Royal Statistical Society* (Series B), 1969, **31**, 486–498. (b)

GOODMAN, L. A. The multivariate analysis of qualitative data: Interactions among multiple classifications. *Journal of the American Statistical Association*, 1970, **65**, 226–256.

GOODMAN, L. A. The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for building models for multiple classifications. *Technometrics*, 1971, **13**, 33–61 .(a)

GOODMAN, L. A. Partitioning of chi-square, analysis of marginal contingency tables, and estimation of expected frequencies in multidimensional contingency tables. *Journal of the American Statistical Association*, 1971, **66**, 339–344. (b)

HABERMAN, S. J. The general log-linear model. Unpublished doctoral dissertation, University of Chicago, 1970.

KU, H. H., VARNER, R. N., & KULLBACK, S. On the analysis of multidimensional contingency tables. *Journal of the American Statistical Association*, 1971, **66**, 55–64.

LANCASTER, H. O. Complex contingency tables treated by the partition of χ^2. *Journal of the Royal Statistical Society* (Series B), 1951, **13**, 242–249.

LINDLEY, D. V. The Bayesian analysis of contingency tables. *Annals of Mathematical Statistics*, 1964, **35**, 1622–1643.

MILLER, R. G., JR. *Simultaneous statistical inference*. New York: McGraw-Hill, 1966.

MOOD, A. M., & GRAYBILL, F. A. *Introduction to the theory of statistics*. (2nd ed.) New York: McGraw-Hill, 1963.

MOSTELLER, F. Association and estimation in contingency tables. *Journal of the American Statistical Association*, 1968, **63**, 1–28.

SCHEFFÉ, H. *The analysis of variance*. New York: Wiley 1959.

SHAFFER, J. P. Exact procedures for the analysis of multidimensional contingency tables. *Behavior Research Methods and Instrumentation*, 1972, **4**, 231–236.

SIMPSON, E. H. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society* (Series B), 1951, **13**, 238–241.

SUTCLIFFE, J. P. A general method of analysis of frequency data for multiple classification designs. *Psychological Bulletin*, 1957, **54**, 134–137.

(Received September 14, 1971)

Psychological Bulletin
1973, Vol. 79, No. 2, 142–144

CORRELATION BETWEEN GENETICALLY MATCHED GROUPS VERSUS RELIABILITY THEORY:

A REPLY TO JENSEN

JOHN K. MILLER
University of Rochester

DAVID LEVINE
University of Nebraska

In a "Note on Why Genetic Correlations Are Not Squared" Jensen draws an analogy to classical theory of test reliability to explain why "geneticists do not square the correlations between twins . . . in order to obtain the percentage of variance explained by genetic factors [p. 223]." It is concluded, in light of theoretical assumptions inherent in the statistical model, that the analogy is improper and that its acceptance leads to overestimation of variance attributable to genetic factors.

Recently Jensen (1971) stated the case for the practice of interpreting correlations between phenotypes (observed measures of heritable traits taken on genetically related subjects) as direct proportional measures of phenotype variance attributable to trait heritability. The issue here is methodological: Is the correlation between phenotypes properly interpreted as the proportion of observed variance attributable to a hypothetical measure called the genotype? Genotypes are defined as the heritable components of traits whose development is subject to both genetic and environmental determinants.

Jensen (1971) contends that $r_{pp} = r^2_{pg}$, where r_{pp} is the correlation between phenotypes (p) and where r_{pg} is the correlation between phenotype and genotype (g), implying that r_{pp} represents the ratio of genotype variance to phenotype variance (S_g^2/S_p^2). For support he appeals to an analogy with measurement theory. Classical test theory defines an individual observed score (X) by the equation $X = t + e$, the sum of a "true" score (t) and an "error" component (e). A priori, t for each person is stable over successive measures of the same trait, while e is random and varies unsystematically from one occasion to the next. Given these assumptions, the proofs are simple and straightforward that both r_{xx} and r_{tx}^2 equal the proportion of observed variance due to "true" variance (S_t^2/S_x^2). Because "errors" are *assumed* to be random, and therefore uncorrelated with "true" scores ($r_{te} = 0$) or with one another ($r_{ee} = 0$), their contribution to test reliability is zero. Jensen treats genetic variance and environmental variance as analogous to true variance and error variance, respectively.

The relation of a phenotypic measure like IQ to its hypothetical components, heritability (g) and environmental factors (E), is governed by the legitimacy of assumptions associated with partitioning p according to the definitional equation $p = g + E$. Environment's contribution to the phenotype must be taken to be a normally distributed random variable such that (*a*) $r_{ge} = 0$; (*b*) $r_{EE} = 0$; (*c*) and the average contribution of E to $p = 0$. The first two assumptions must be satisfied if genotype and environmental variance are not to become confounded in r_{pp}. The third assumption must be satisfied if the confidence interval embracing with known probability an individual g is to be estimated. Failure of theoretical formulations to reflect these statistical assumptions constitutes an irreconcilable conflict that can be resolved only by altering the theory or by finding a more adequate means of statistically relating empirical observations to theoretical constructs. Let us review briefly the conformity of theoretical constructs proposed by Jensen (1968) to the statistical assumptions implicit in the theorem $r_{pp} = r_{pg}^2 = S_g^2/S_p^2$.

1. $r_{ge} = 0$: In the absence of experimental control of environmental factors, psychological measurement of genetic effects on human behavior is limited to the framework of causal comparative studies. In such studies, random assignment of members from the same kinship pair to different groups prior to statistical analysis is possible and desirable. This procedure, however, contributes to controlling only the effects of sampling bias attrib-

utable to differences *within* pairs which could contaminate correlation between matched groups. It would most assuredly have no neutralizing effect upon covariance attributable to similar and systematic environmental influences upon phenotype performance by both members of the same kinship pair. As a consequence, the door remains open to a systematic contribution of environmental factors to phenotype differences *between pairs* irrespective of the groups to which the members of each pair are assigned. Jensen (1968) in fact, has suggested elsewhere that there is "... probably a positive correlation between the quality of environment and potential for intelligence [p. 7]" within any population sampled. It is precisely such environmental similarities to which Jensen (1968, p. 7) appeals to explain agreement in findings among studies based upon different populations. Postulating the existence of such a relationship contradicts the statistical assumption required to permit the interpretation of r_{pp} as a ratio of genetic variance to observed variance (S_g^2/S_p^2). Relaxing the assumption that $r_{gE} = 0$, requires $S_p^2 = S_g^2 + S_E^2 + 2rS_gS_E$, so that r_{pp} exceeds S_g^2/S_p^2 by at least $2rS_gS_E/S_p^2$. The net effect will be the confounding of purely genetic components and the interaction between genetic and environmental factors in r_{pp}.

2. $r_{EE} = 0$: The failure of the assumption of independence between genotype and environment implicitly weakens any assumption of independence between environments (even separate environments) experienced by the two members of a kinship pair. Furthermore, random separation of kinship pairs into different groups could no more efficaciously negate environmentally determined phenotype similarities within pairs than it would control correlation between genetic and environmental factors. Even if, contrary to Jensen's (1968, p. 7) suggestion, r_{gE} *were* zero, the likelihood remains that notable environmental similarities would exist even for kinship pairs reared apart. Prenatal environments of identical twins are not random. Nor are the neonatal environments of twins random, unless separation occurs at birth. Nor, does it appear plausible to assume that twins separated at birth have equal and unrelated prospects for rearing in an enriched or an impoverished environment. The relation between factors like socioeconomic status and measured intelligence even among genetically unrelated members of the same social class is accepted as real and substantial (Jensen, 1968, pp. 15–18). Environmental factors common to genetically related individuals, therefore, could contribute, independent of genotype, to reliable differences between kinship pairs. The result would be a contribution to the correlation between phenotypes (r_{pp}) quite independent of the phenotype–genotype correlation (r_{pg}). Maxwell (1968) has, for example, shown that in the estimation of reliability, relaxing the assumption of uncorrelated errors (the assumption of uncorrelated environments in the genetic analogy) will result in a biased estimate of the reliability coefficient (r_{pp} in the genetic analogy). Such bias is usually in the direction of overestimation.

3. The average contribution of E to p is 0 (i.e., Mean of $E = 0$): The final assumption required to satisfy fully an analogy of a reliability coefficient with correlation between phenotypes does not directly affect the interpretation of r_{pg} as a function of r_{pp}. Yet failure to fulfill the assumption has important consequences for the practical utility of r_{pp} in relation to trait heritability. Environment is viewed as a "threshold variable" whose effect on phenotype performance is limited to the extent to which it hinders or facilitates the attainment of full genetic potential (Jensen, 1968, pp. 10–14). This precludes estimation of genetic analogues to the standard error of measurement and the interval estimation of true score performance of single individuals. As a threshold variable, environment must be conceived of as essentially negative, having a negative mean and a maximum value of zero. The mean of observed performance, therefore, cannot be equated with the mean of genotype performance as observed and true means may be equated in reliability theory. And, as a result, a "standard error of genetic potential" cannot be estimated as a direct function of phenotype variability.

With due deference to the scholarship embodied in Jensen's attack on the problem of psychological heritability, his faith in the potency of correlation between genetically matched groups as the genetic proportion of observed variance is misplaced.

REFERENCES

JENSEN, A. R. Social class, race, and genetics. *American Educational Research Journal*, 1968, **5**, 1–42.

JENSEN, A. R. Note on why genetic correlations are not squared. *Psychological Bulletin*, 1971, **75**, 223–224.

MAXWELL, A. E. The effect of correlated errors on estimates of reliability coefficients. *Educational and Psychological Measurement*, 1968, **28**, 803–811.

(Received September 23, 1971)

Manuscripts Accepted for Publication in the *Psychological Bulletin*

Neglected Developmental Variables in the Study of Species Identification in Birds. Gilbert Gottlieb (Psychology Laboratory, Dorothea Dix Hospital, Raleigh, North Carolina 27611).

Long-Term Psychological Effects of LSD. Spencer A. McWilliams (Department of Psychology, University of Arizona, Tucson, Arizona 85721).

Representation of Frequency in Memory. William C. Howell (Department of Psychology, Rice University, Houston, Texas 77001).

Theories of Dream Function. Janet O. Dallett (929 Gretna Green Way, Los Angeles, California 90049).

An Analysis of Responses to Adjectives: A Reply to Samelson. P. M. Bentler (Department of Psychology, University of California, Los Angles, California 90024).

VOL. 79, NO. 3 MARCH 1973

Psychological Bulletin

Copyright © 1973 by the American Psychological Association, Inc.

SELF-CONTROL OF CARDIAC FUNCTIONING:
A PROMISE AS YET UNFULFILLED[1]

EDWARD B. BLANCHARD[2] AND LARRY D. YOUNG

University of Mississippi Medical Center

The research in the operant conditioning or self-control of four cardiac functions, heart rate level, heart rate variability, blood pressure, and cardiac arrhythmias, is reviewed. Although the fact of the self-control of these functions seems well established, with few exceptions the magnitude of change is statistically significant rather than clinically significant. A few studies, all involving numerous training sessions, have reported large magnitude changes. Several methodological issues are raised and discussed. Finally, alternative modes of achieving self-control of cardiac function, such as progressive relaxation and Yoga exercises are discussed.

Since the mid-1960s an increasing number of articles on the topic of the operant conditioning and/or establishment of self-control of cardiovascular processes in humans have appeared in the literature. These articles hold out the promise, either implicitly or explicitly, of being able to treat cardiovascular disorders by psychological, rather than pharmacological, means. However, a critical examination of this literature, which is the principal purpose of the present article, reveals that, for the most part, this promise is, as yet, unfulfilled. With rare exception, the results are reported for changes in cardiovascular functioning of normal subjects instead of patients and the magnitude of change reported has been of a *statistically* significant, rather than *clinically* significant, magnitude.

The research is summarized in terms of the four cardiovascular responses studied, with particular emphasis on the magnitude of change obtained. From this summarization, several methodological issues become apparent and are subsequently discussed.

HEART RATE

In Table 1 are summarized the studies involving change in heart rate as the dependent variable. Omitted from this table, and from intensive consideration by this review, are studies by Shearn (1962), Frazier (1966), and Brener (1966), all of which used a shock avoidance procedure. Basically these studies allowed the subject to avoid mild but painful electrical shocks by emitting a heart rate of the appropriate magnitude. These studies, which did show that subjects could learn some degree of heart rate control, are omitted because such a paradigm was not thought to be relevant for clinical practice.

Inspection of Table 1 reveals that each experimenter has tended to use his own individual method of determining if subjects can change their heart rate. Although, with three exceptions, the magnitude of the change in heart rate achieved is quite small, ranging 1–7 beats per minute, nevertheless, it seems fairly well established that small, but statistically significant, changes in heart rate can

---

[1] Preparation of this manuscript was supported in part by Grant 1R01HL14906-01 from the National Heart and Lung Institute, Robert W. Scott, principal investigator. The authors wish to thank W. Stewart Agras and Mischel Hersen for their critical comments on earlier versions of this manuscript.

[2] Requests for reprints should be sent to Edward B. Blanchard, Psychiatry Department, University of Mississippi Medical Center, Jackson, Mississippi 39216.

705(5)
4/1/74

TABLE 1

Summary of Self-Control of Heart Rate

Authors	Independent variables								Results
	Desired response	Reinforcement	Feedback type	Subjects informed of response	Trial length	Trials per session	Number of sessions	Controls	Magnitude of heart rate change
Engel & Hansen, 1966	Decrease HR	½¢ per second of correct response	Binary visual	No	25 minutes	1	6	Yoked	E: −.5 BPM (range: 2.6 to −2.6 BPM) C: 5.6 BPM
Engel & Chism, 1967	Increase HR	½¢ per second of correct response	Binary visual	No	25 minutes	1	6	Yoked	E: 5.9 BPM (range: 4.2 to 9.8 BPM) C: 3.0 BPM
Levene, Engel, & Pearson, 1968	Alternately increase and decrease HR	None	Binary visual	No	1 minute	25	Varied from 2–4 (until response was learned)	None	E: 4.8 BPM on increase trials (2 of 5 subjects showed consistent alternation) −2.5 BPM on decrease trials
Brener & Hothersall, 1966	Alternately increase and decrease HR	None	Binary auditory	No	50 IBIs (50 seconds)	8	1	None	E: HR > base line on 85% of beats on increase trials. HR < base line on 55% of beats on decrease trials (HR < base line on 80% of beats n post decrease trial)

Note. Abbreviations: E = experimental, C = control, BPM = beats per minute, HR = heart rate, IBI = interbeat interval.

Table 1—*(Continued)*

Authors	Independent variables								Results
	Desired response	Reinforcement	Feedback type	Subjects informed of response	Trial length	Trials per session	Number of sessions	Controls	Magnitude of heart rate change
Brener & Hothersall, 1967	Alternately increase and decrease HR	None	Binary auditory	No	50 IBIs (50 seconds)	22	2	None	E: 8–10 BPM difference between increase and decrease trials
Brener, Kleinman & Goesling, 1969	Alternately increase and decrease HR	None	None	Yes	50 IBIs (50 seconds)	29	1	None	E: 1–3 BPM difference between increase and decrease trials
	Alternately increase and decrease HR without feedback	None	Binary auditory for 100%, 50%, and 0% of trials	Yes	50 IBIs (50 seconds)	18	2	None	100% feedback significantly better (by 2–3 BPM difference between increase and decrease trials) than 0% feedback; 100% feedback showed 10 BPM difference at second trial; second session performance significantly better than first session
Johns, 1970	Increase and decrease HR	None	Binary auditory	E: Yes	70 seconds	19	1		E: raise: 4.6 BPM; lower: no change

Table 1—(*Continued*)

Authors	Independent variables								Results
	Desired response	Reinforcement	Feedback type	Subjects informed of response	Trial length	Trials per session	Number of sessions	Controls	Magnitude of heart rate change
				C1: Yes C2: No C3: No				C1: Information no feedback. C2: No information-feedback. C3: No information-no feedback	C1: raise: no change; lower: −2.3 BPM C2: no change C3: no change
Finley, 1971	Decrease HR	None	Proportional visual	No	5 minutes	11	1	Inaccurate feedback	E: −1.7 BPM feedback versus no feedback E versus C: −1.5 BPM
Sroufe, 1971	Increase HR	None	Proportional visual	Yes	5 minutes	3	1	None	E: 10 BPM for shallow versus deep respiratory volume
Peters, Scott, & Gillespie, 1971	Increase HR Decrease HR	Contingent commercial television	None	No	20 minutes	1	13 10	Single subje Single subject	19 BPM −19 BPM
Headrick, Feather, & Wells, 1971	Alternately increase and decrease HR	None	Proportional auditory	No	1 minute	10	1	None	E: raise: 3.3 BPM; lower: no change
	Increase and decrease HR	None	Proportional visual	Yes	1 minute		12	Single subject	E: raise: 30 BPM; lower: −5 BPM
Bergman & Johnson, 1971	Increase and decrease HR	None	None	Only experimental subjects	6 IBIs (5 seconds)	30	1	Uninformed subjects	raise: E>C, +1.5 BPM lower: no difference E versus C

Table 1—(Continued)

Authors	Independent variables								Results
	Desired response	Reinforcement	Feedback type	Subjects informed of response	Trial length	Trials per session	Number of sessions	Controls	Magnitude of heart rate change
Blanchard, Young & McLeod, 1972	Increase and decrease HR	None	Proportional visual	Yes	1 minute	10	1	None	Low aware subjects could raise (2.9 BPM) and lower (−2.8 BPM) HR from own base line
Blanchard & Young 1972	Increase and decrease HR	None	Proportional visual and proportional auditory	Yes	1 minute	8	2	No feedback	E combined: raise: +4.2 BPM; C .3 BPM; lower: −3.3 BPM C −2.2 BPM
Scott, Blanchard, Edmunson, & Young, 1972	Increase HR	Contingent commercial television or 1¢/10 second correctr esponse	Binary visual	No, then yes	20 minutes	1	34	Single subject	16 BPM with contingent money
	Decrease HR			Yes			52		−16 BPM, decided clinical improvement
	Increase HR			No			18		30 BPM
Stephens, Harris, & Brady, 1972	Increase and decrease HR	$\frac{1}{2}$¢/second of correct response	Proportional visual and proportional auditory	Yes	5 or 30 minutes	8 or 1	17	Single subject	Increase = 23 BPM decrease = 5 BPM
					5 or 60 minutes	4 or 1	18		Increase = 24 BPM
					5 minutes	8	10		Increase = 15 BPM decrease = 6 BPM
					5 or 60 minutes	8 or 1	12		Increase = 29 BPM

Table 1—(*Continued*)

Authors	Independent variables								Results
	Desired response	Reinforcement	Feedback type	Subjects informed of response	Trial length	Trials per session	Number of sessions	Controls	Magnitude of heart rate change
Bergman & Johnson, 1972	Increase HR	None	Binary visual for half of subjects in all groups E_1 = none E_2 = heart sounds	E_1 and E_2 = yes C = no	20 seconds	30	1	Uninformed, no feedback	Binary visual feedback ineffective as reinforcer; informed, instructed subjects E_1 & $E_2 > C$: 1.8 BPM $E_1 = E_2$

be obtained through giving subjects feedback or reinforcement. However, these changes, though of sufficient magnitude to be statistically significant, are not clinically significant. For the purposes of this review, a change in response is considered to be clinically significant if it is equal to either 20% of the base rate or, in the case of an abnormally high response, the change is such as to return the response to the normal range.

The three studies which have demonstrated large magnitude changes in heart rate include a single subject whom Headrick, Feather, and Wells (1971) ran for 14 sessions and who demonstrated within-trial increases in heart rate of 30 beats per minute for the short 1-minute trial. Recently Stephens, Harris, and Brady (1972) have reported on four normal subjects who were run under a complex combination of feedback and reinforcement conditions for 10–18 sessions of trials varying 5–60 minutes. In all four subjects clinically significant within-trial increases in heart rate were obtained, ranging 15–29 beats per minute; moreover, two of the subjects were able to decrease their heart rate 5–8 beats per minute from base line. A final study by Scott et al. (1972)[3] reported clinically significant increases in heart rate of 16 and 25 beats per minute in two normal subjects which were maintained over three consecutive daily sessions. More importantly, a stable (over daily sessions) 16 beats per minute decrease in heart rate was obtained in a patient suffering from tachycardia. Coincident with the decrease in heart rate (in the latter subject) were several clinical improvements including obtaining gainful employment. This last case is the only reported direct clinical application of the studies of heart rate conditioning to date.

The methodological issues and conflicts which have arisen in this area are presented next.

Operant Conditioning versus Self-Control

The first issue is one of terminology, yet it extends beyond mere semantics. Engel and his associates refer to their operations as operant conditioning. In addition to visual feedback, their subjects were given small sums of money for accumulated experimental time during which their heart rate was in the appropriate direction relative to a criterion heart rate. Stephens et al. (1972) followed a similar procedure.

---

[3] R. W. Scott, E. B. Blanchard, E. Edmundson, & L. D. Young. An improved shaping procedure for clinical cardiac control. Paper presented at Southeastern Psychological Association meeting, Atlanta, 1972.

Most of the other studies of heart rate change are described in terms of self-control established through external sensory feedback. It is possible, of course, to construe all of these studies as examples of operant conditioning in which the reinforcement is knowledge of having successfully emitted the correct response. On the other hand, it is equally possible to construe the reinforcement, especially that used by Bergman and Johnson (1972), in informational terms as binary feedback, since subjects receive reinforcement if they are correct and do not receive it if they are not making the correct response.

If one considers knowledge of results to be a reinforcer, feedback and reinforcement are inextricably confounded in the feedback studies. Therefore, it is not possible to settle the question by merely giving one group of subjects feedback and a second group reinforcement, and then comparing their performance. However, it would be possible to determine if the reinforcement used by Engel has an effect above that achieved by feedback by comparing the performance of groups of subjects run in Engel's paradigm with and without monetary reinforcement.

Type of Feedback

There are two dimensions in the classification of types of feedback used. The first is sensory modality. Feedback has been given to subjects by either auditory or visual means, or by both, with no apparent rationale given for preferring one mode to the other. Blanchard and Young (1972), in their study of the relative efficacy of these two modes, made both within- and between-subject comparisons, and found no statistical advantage for one sensory modality over the other. Thus it would seem that the decision as to which sensory modality one should use, particularly in a clinical procedure, must rest on factors other than efficacy.

In describing the second classification dimension, we have introduced some new terminology which may require explanation. By *binary* feedback we refer to a situation such as that in studies by Engel and his associates and by Brener and his associates wherein the subject's heart rate in terms of interbeat intervals, the reciprocal of heart rate, is compared to some criterion on a beat-by-beat basis. After each comparison, the subject is automatically informed as to whether or not he has met or exceeded the criterion by the presence or absence of the feedback signal, respectively. This class of feedback has been delivered in either the auditory mode (Brener, Kleinman, & Goesling, 1969) or the visual mode (Engel & Chism, 1967b; Engel & Hansen, 1966).

By *proportional* feedback, we refer to a situation such as that in studies by Finley (1971), Sroufe (1971), Blanchard et al. (1972) in which the subject is provided with direct knowledge of heart rate on a beat-by-beat basis. With this kind of feedback, the subject is told not only if he is right or wrong as in binary feedback, but also by how much. At the time of this review no one has compared binary and proportional feedback with regard to relative efficacy.

Knowledge of Response by the Subject

Whether or not the subjects should be informed of the particular response which they are to change is an issue in the field of heart rate control which provokes almost as much mixed opinion as the particular feedback arrangement to use. From Table 1 one can see that researchers are fairly well divided as to whether or not to inform the subjects of the nature of the response. For example, Brener changed his policy on informing the subjects: in his first two studies (Brener & Hothersall, 1966, 1967) the subjects were not informed; in his later study (Brener et al., 1969) they were informed.

This difference of opinion seems to date back to the early work of Engel and his associates. For reasons which are unclear in their reports, they elected not to inform subjects what response was related to the feedback and reinforcement. Engel and Hansen (1966) found, post hoc, that subjects who did not correctly infer the response that was related to the feedback and reinforcement contingency showed evidence of learning the heart-rate-slowing response whereas four of their five "nonlearners" correctly guessed that heart rate was the response. In a second study (Engel & Chism, 1967b) although all five subjects showed evidence of heart rate speeding, four of the five subjects did not know the nature of the response.

Bergman and Johnson (1972) have taken a step toward resolving this issue by treating awareness of the response as an independent variable. Of two groups of subjects receiving no feedback of heart rate, those informed of the response raised it significantly higher ($\Delta\bar{X} = 1.8$ beats per minute) than uninformed subjects trying to control an unspecified response. Interestingly, a third group of subjects who were both informed and given feedback in the form of heart sounds did no better than the informed, no-feedback subjects.

One would hope that further research in this area will show that knowledge of the response being changed, at the very least, has no detrimental effect since in a clinical application of these procedures one would prefer to be able to inform patients fully about the procedure.

Types of Control Groups

Several different policies on control groups have been followed in these studies. One policy has been to use no formal control group, but instead to use subjects as their own controls by comparing heart rate during experimental trials to that recorded during some other period. This within-subject control policy was followed by Blanchard, Young, and McLeod, 1972; Brener and Hothersall, 1966, 1967; Brener et al., 1969; Headrick et al., 1971; Levene, Engel, and Pearson, 1968; and Stephens et al., 1972.

A second policy on control groups, which has been followed by Engel and Hansen (1966) and Engel and Chism (1967b), is the use of yoked controls. In this procedure the control subject receives the same feedback signal as the experimental subject to whom he is yoked. A similar form of false feedback has been used by Finley (1971).

Church (1964) has pointed out the statistical and methodological fallacies inherent in the use of yoked controls. Examination of the results for subjects given some form of false feedback reveals that the performance of these yoked control subjects tends to deteriorate, that is, their responses are in the wrong direction. For example, in the study of heart-rate-slowing by Engel and Hansen (1966) the control subjects achieved a mean increase in heart rate of 5.6 beats per minute. Furthermore, Finley's (1971) control subjects showed an increased heart rate ($\Delta\bar{X} = .8$ beats per minute) when they received false feedback as contrasted to times when they received no feedback. Comparisons of changes in heart rate for experimental and control subjects thus yield spuriously large effects because of capitalizing on differences which include the deterioration of the controls.

The appropriate control condition would be one in which subjects receive the same instructions as the experimental subjects but no feedback. The inclusion of an instructed, no-feedback control has been followed by Blanchard and Young (1972), Bergman and Johnson (1972), Brener et al. (1969), Johns (1970), and Sroufe (1971).

Changes in Heart Rate in the Absence of Feedback or Reinforcement

The use of a no-feedback condition enables one to detect if subjects can significantly change their heart rate in the absence of external feedback or reinforcement. The results on this point are mixed. Bergman and Johnson (1971) found both significant increases and decreases for very brief (4–6 seconds) trials in the absence of any feedback. In a later study, described above, Bergman and Johnson (1972) found that subjects instructed to increase heart rate but given no feedback did as well as similarly instructed subjects who received a form of auditory feedback.

Although the subjects of Brener et al. (1969) who received feedback on 100% of the trials showed significantly greater heart rate control than those subjects who received no feedback, the latter group did show a small ($\Delta\bar{X} = 2.4$ beats per minute) but significant difference between raise and lower trials. However, since the results were reported as differences, no directional statement can be made from their data. Both Johns (1970) and Blanchard and Young (1972) found that their no-feedback control subjects could lower heart rate but could not raise it ($\bar{X} = -2.3$ beats per minute, $\bar{X} = -2.2$ beats per minute, respectively). At this point one cannot conclude definitely that feedback enables subjects to change heart rate in either direction more than subjects who are similarly instructed but receive no feedback.

Another way of viewing this question is to examine how well subjects who have been trained to control heart rate with feedback can control their heart rate in the absence of feedback. This is a point of considerable clinical significance since transfer of the self-control that patients learn in an experimental setting to their natural environment is essential. Brener et al. (1969) found significantly greater generalized self-control for subjects given feedback on 100% of trials than to subjects given no feedback.

Sroufe (1971), working primarily with heart rate variability (to be discussed later), has achieved similar results. His subjects, who were taught to control heart rate variability through paced respiration and who were given proportional visual feedback, showed significant transfer of self-control to a no-feedback state after only a brief (5-minute) training period.

Respiratory Effects

An issue in the field of self-control of cardiac function surrounded by much controversy is the relative importance of mediational variables. Although mediation by means of cognitive or unspecified skeletal muscular mechanisms has been suggested, the primary focus of the mediation controversy has been on the effects of respiratory variables. Experimental procedures designed to answer whether or not the obtained heart rate change resulted from respiratory changes fall into four general categories. Although no investigator failed to monitor respiration, some did little else. Brener and Hothersall (1966) and Brener et al. (1969) report monitoring respiration but do not mention any instructions concerning breathing nor do they report any analyses of the observed differences.

Blanchard and his associates (Blanchard et al., 1972; Blanchard & Young, 1972) and Engel and his associates (Engel & Chism, 1967b; Engel & Hansen, 1966; Levene et al., 1968) gave all subjects in their investigations the explicit instructions not to change their breathing pattern from normal. Again, however, only monitoring and visual inspection of the data were used to reach the conclusions that respiration did not change significantly and that the observed heart rate differences did not result from respiratory changes. Levene et al. (1968) did carry the investigation a step further, however, by teaching two subjects whose respiration rates were variable to duplicate their earlier respiration rate through a pacing technique. This resulted in changes in heart rate variability but not in overall heart rate.

The third approach to the problem, namely, manipulation of respiration rate to insure its constancy during heart rate conditioning sessions, was employed by Brener and Hothersall (1967), Johns (1970), and Sroufe (1969) and basically consists of having subjects pace their respiration to a metronome or rhythmic puff of air set at their adaptation respiration rate. Even with respiration rate held constant experimentally, subjects were still able to alter their heart rate significantly in the appropriate direction. However, as Sroufe pointed out, ability to control heart rate appears to deteriorate with any session when the subject is attempting to control two systems simultaneously.

The final method used (Finley, 1971; Headrick et al., 1971; Hnatiow & Lang, 1965; Lang, Sroufe, & Hastings, 1967) to insure that changes in heart rate were not the result of respiratory differences is to monitor, analyze, and control statistically through covariance analyses any systematic variations in respiration that are found. There have been no significant relationships between heart rate change and respiratory changes with two minor exceptions: Headrick et al. (1971) found a significant correlation between breathing amplitude and heart rate for only one subject; and Hnatiow and Lang's (1965) single significant correlation was between average heart rate deviation during the final no-feedback period and the average inspiration to total cycle ratio for the final feedback period.

A summary of cardiorespiratory relations is not complete without the inclusion of two studies which focus specifically on them. In the earlier investigation (Engel & Chism, 1967a) 20% changes in respiration rate sustained for a 10-minute period did not cause changes in average heart rate. However, an inverse relation was noted between respiration rate and heart rate variability. Sroufe's (1971) series of three experiments was broader in scope since he manipulated respiratory volume in addition to respiration rate. His results confirmed the earlier finding of Engel and Chism (1967a) that changes in respiration rate affect only heart rate variability with increases in respiration rate producing a more stable heart rate. However, changes in respiratory volume affected both the level and variability of heart rate with deeper breathing producing a higher but more variable heart rate.

In light of the evidence, monitoring of respiratory variables seems mandatory for those investigations attempting to show heart rate control. However, monitoring alone is not sufficient; the effects of respiratory changes must be controlled either experimentally or statistically to eliminate this possible source of confounding. As Sroufe (1971) has so aptly observed, "The fact that cardiac control is possible with respiration experimentally controlled merely suggests that subjects *can* control rate independent of respiration, not that respiration is of minor importance [p. 654]."

Sroufe's (1971) research carries several implications and suggestions for subsequent research in the area of heart rate self-control, particularly clinical studies: Since respiration and skeletal muscle activity are more discriminable than cardiac activity, they might profitably be used early in training and gradually faded as the desired response is acquired. As Sroufe (1971) showed, utilization of these or similar techniques improved the transfer of the learned cardiac control from periods where feedback is present to periods where it is absent; ultimately, utilization improved the first step in establishing generalization from the experimental chamber to the natural environment.

HEART RATE VARIABILITY

Lang and his associates have been the principal researchers investigating the self-control of heart rate variability. All of their studies have used the same basic experimental procedure: The experimental subjects are given proportional visual feedback of their heart rate and asked to keep it within ±3 beats per minute of a set point (their mean preexperimental heart rate). Dependent measures are the standard deviation of the distribution of interbeat intervals and time on target (within the ± beats per minute range).

In the first study (Hnatiow & Lang, 1965) experimental subjects performed significantly better than yoked controls who received false feedback and who were told only to track the visual display, with no mention of controlling it.

The second study (Lang, Sroufe, & Hastings, 1967) was a refinement of the first to include additional controls. Subjects in the experimental group again received proportional visual feedback of heart rate, were told what the feedback display represented, and were instructed to try to keep their heart rate within a 6 beat per minute range. Those in one control group received the same instructions but the feedback display of their yoked experimental partner. The other two control groups were not informed concerning what the visual display represented, and were told only to track the visual display.

The experimental group was significantly better at reducing heart rate variability than any of the controls. Of special note are the within-subject analyses which showed that the experimental subjects improved significantly from no-feedback to feedback-available conditions. The control subjects yoked to the experimental subjects, who were also supposed to control heart rate, did poorest, supporting our previous criticism of this control procedure.

In the third study of this series (Sroufe, 1969), subjects again were given either true or false proportional visual feedback of heart rate and told to keep it within a specified range. The purpose of the study was to control for respiration effects and the possibility that subjects were reducing heart rate variability through respiratory means. Thus three groups were taught, over several sessions, to breathe at a fixed rate, while a fourth group had no restrictions placed on their breathing. All subjects received two training sessions before data were collected in a third session of 1-minute trials. Again, subjects given true feedback showed significant reductions in heart rate variability during times it was available compared both to no-feedback periods and to subjects receiving false feedback. For subjects breathing at fixed rates, performance tended to deteriorate over time while no such deterioration was found for subjects not under respiratory restriction.

Interestingly, a final study in this series (Hnatiow, 1971) failed to replicate the early findings of better control by subjects receiving feedback than the yoked controls in spite of using identical equipment and procedures.

To date, Lang and his associates have reported no clinical application of their procedure. In fact, all of their work has been with normal college students.

BLOOD PRESSURE

In Table 2 are summarized the studies involving change in blood pressure as the dependent variable.

In all of their blood pressure control studies, Shapiro and his colleagues at Harvard have used the same basic experimental procedure: A subject receives binary feedback, both visual and auditory, as to whether his blood pressure is at or in the appropriate direction away from a criterion blood pressure value on each heart beat; in addition to this sensory feedback, the subject receives a reinforcement of a brief presentation of a slide of either a nude, a landscape, or an indication of small financial reward on some fixed-ratio schedule for a blood pressure response in the appropriate direction.

The method of delivering this feedback involves inflating an arm cuff to a pressure very close to either systolic or diastolic blood pressure, depending on which is of interest, then varying the pressure in small increments and noting the presence or disappearance of Korotkoff sounds within a specified time interval after each heart beat. Feedback delivery is dependent upon the appropriate response. For example, in trying to shape a decrease in systolic blood pressure, feedback would be given the subject if no Korotkoff sound was detected following the R wave of the electrocardiogram. Had a sound been detected, indicating that the systolic pressure was above the cuff pressure, no feedback would be given. Typically subjects are given 25 trials per session with a trial being 50 interbeat intervals long.

As can be seen in Table 2, with the exception of their recently completed clinical trial (Benson, Shapiro, Tursky, & Schwartz, 1971), the magnitude of change in blood pressure has been rather small, averaging less than 1 millimeter of mercrury per trial for subjects who were increasing blood pressure and about 2–3 millimeters of mercury per trial for subjects lowering their blood pressure. Using as an index of the efficacy of their procedure the average blood pressure on the final block of five trials, the average amount of increase in blood pressure has been 1–2 millimeters of mercury while the average decrease in subjects lowering blood pressure has been 4–6 millimeters of mercury.

Despite their impressive series of studies, many of the criticisms previously leveled at the research on self-control of heart rate can be supplied to the blood pressure self-control studies. First and foremost, it is not possible to tell what are the active ingredients in the Shapiro procedure since subjects receive both binary auditory and visual feedback as well as reinforcement on some fixed ratio schedule. Whether reinforcement contributes to the efficacy of the procedure is not known, nor are the relative efficacies of the types of feedback.

A second problem with these studies has been the failure to include a no-feedback control in order to evaluate the efficacy of their combination of procedures over merely instructing subjects to lower their blood pressure. In one study (Shapiro et al., 1970a) a group reinforced on a random basis was included. Results showed that groups reinforced for raising or lowering blood pressures, while different from each other, did not differ from this group which was reinforced randomly.

TABLE 2

Summary of Self-Control of Blood Pressure

Authors	Independent variables								Results
	Desired response	Reinforcement	Feedback	Subjects informed of response	Trial length	Trials per session	Number of sessions	Controls	Results: magnitude of blood pressure change
Shapiro, Tursky, Gershon, & Stern, 1969	E_1 = Increase systolic BP E_2 = Decrease systolic BP	Slides of nudes FR 20	Binary auditory & visual	No	65 seconds	25	1	None	E_1: −.8 millimeters Hg $E_1 > E_2$ E_2: −4.8 millimeters Hg
Shapiro, Tursky, & Schwartz, 1970a	E_1 = Increase systolic BP E_2 = Decrease systolic BP	Slides (1/3 nudes, 1/3 landscapes, 1/3 money... 10–25¢) FR 20	Binary auditory & visual	No	50 IBI	25	2	Random reinforcement (50% probability for each IBI)	1. E_1: 2.3 millimeters Hg; C: 0 millimeters Hg $E_1 > C = E_2$
	E_1, C, E_2 = Decrease systolic BP								2. E_1: −2.0 millimeters Hg; C: −2.0 millimeters Hg; E_2: −4.5 millimeters Hg. $E_1 = C > E_2$
Shapiro, Tursky, & Schwartz, 1970b	E_1 = Increase HR E_2 = Decrease HR	Slides of nudes FR 20	Binary visual for all subjects plus half received binary auditory & other half proportional auditory	Yes	50 IBI	25	1	None	E_1: HR = .8 BPM; BP = .5 millimeter Hg E_2: HR − 5.0 BPM; BP = −.5 millimeter Hg For BP $E_1 = E_2$
Schwartz, 1972	HR-BP Integration & differentiation	Slides (landscapes, nudes, money) FR 12	Binary auditory & visual	No	50 IBI	25	1	None	HR = BP Integration easier to achieve than HR = BP differentiation

Note. Abbreviations: E = experimental, Hg = mercury, IBI = interbeat interval, HR = heart rate, BP = blood pressure, FR = fixed ratio.

Table 2—*(Continued)*

Authors	Independent variables								Results
	Desired response	Reinforcement	Feedback	Subjects informed of response	Trial length	Trials per session	Number of sessions	Controls	Results: magnitude of blood pressure change
Shapiro, Schwartz, & Tursky, 1972	E_1 = Increase diastolic BP E_2 = Decrease diastolic BP	Slides (landscapes, nudes, money) FR 20	Binary auditory & visual	No	50 IBI	35	1	None	E_1: 7/10 increased BP, $\bar{x} = 4.0$ mm. voluntary control maintained in extinction E_2: 8/10 decreased BP $\bar{x} = -2.0$ mm.
Schwartz, Shapiro, & Tursky, 1971	E_1 = Increase HR & Systolic BP E_2 = Decrease HR & Systolic BP	Slides (landscapes, nudes, money) FR 12	Binary auditory & visual	No	50 IBI	40	1	None	E_1: HR = 1.2 BPM; BP = 2.2 millimeters Hg E_2: HR = −6.0 BPM; BP = −4.8 millimeters Hg $E_1 > E_2$
Benson, Shapiro, Tursky, & Schwartz, 1971	Decrease systolic BP	Slides (landscapes, nudes, money) 5¢/slide FR 20	Binary auditory & visual	Yes	50 IBI	30	8–34 conditioning sessions $\bar{x} = 21.7$	None	$\bar{x} = -16.5$ millimeters Hg (Range: =.9 to −33.8 millimeters Hg)
Brener & Kleinman, 1970	Decrease systolic BP	None	Proportional visual	Yes	50 seconds	20	2	To pay attention to feedback display	E (Session 1): −16 millimeter Hg; (Session 2): −10 millimeter Hg. C (Session 1): No change; (Session 2): No change
Hnatiow, 1971	Decrease systolic BP variability	None	Proportional visual	E Yes C No	5 minutes	3	1	Track display of yoked partner	E = C on blood pressure variability reduction E reduced blood pressure more than C (but this was not the task)

Thus to date, this group of investigators has shown primarily that subjects reinforced for raising blood pressure are different from subjects reinforced for lowering blood pressure.

Despite the methodological problems with these studies, the Shapiro group certainly deserves praise for their pioneering work in this area. Recently they have focused on self-control of diastolic blood pressure (Shapiro et al., 1972) and the degree of integration and differentiation of the cardiovascular system (Schwartz, 1972; Schwartz, Shapiro, & Tursky, 1971). The former line of research, using the same basic experimental paradigm, has shown that diastolic blood pressure can be altered significantly by this procedure and, of much interest from a clinical point of view, that the changes show some degree of stability after the feedback and reinforcement are no longer available. In the latter research area, a fair degree of independence of heart rate and blood pressure have been shown.

The study by Benson et al. (1971) answers one of our major criticisms leveled at the blood pressure self-control research, namely, the exclusive reliance on normal subjects. Benson applied Shapiro's basic blood pressure control procedure to a group of seven hypertensive patients. Their average systolic blood pressure during the final base-line periods was 165 millimeters of mercury. Clinically significant decreases in blood pressure were obtained in five of the seven patients ($\bar{X} =$ 22.6 millimeters of mercury).

As with the heart rate studies which showed large scale changes in that response, numerous trials were necessary to obtain this magnitude of change ($\bar{X} = 21.7$ conditioning sessions of 25 trials each). While no data are available on the stability of the change when the feedback was no longer available, or on the degree of transfer of the decreased blood pressure outside of the laboratory, these results are provocative enough to warrant following up and certainly to increase the clinical promise held out by this research.

As can be seen in Table 2 there has been only one important study of blood pressure self-control conducted by an investigator outside of the Shapiro group at Harvard, and that is by Brener and Kleinman (1970). This study, which utilized a form of proportional visual feedback, has several commendable features: The size of the change in blood pressure, approximately 15 millimeters of mercury in the first session and approximately 12 millimeters of mercury in the second, is considerably larger than that reported by Shapiro and his colleagues with normal subjects, and is of a magnitude which could be clinically significant. Second, control subjects who received the same type of feedback of their own blood pressure but were not instructed to change their blood pressure were used and showed no overall change in blood pressure. A further control group is needed, however, in which subjects, who are instructed to try to lower blood pressure, receive no feedback.

The form of feedback used within the classification system presented previously is closest to proportional visual feedback. Subjects were able to view a manometer which continuously displayed pressure in the occluding staff. The cuff pressure, and hence the feedback received by subjects, oscillated about the individual subject's systolic blood pressure. An interesting aspect of this procedure was the taking of blood pressure on the left index finger rather than upper arm. Such a procedure minimizes the problem ischaemia.

The size changes in blood pressure obtained by Brener and Kleinman's procedure are much larger than those obtained by Shapiro and his colleagues in a similar length of time on a similar population of normotensive college-age males. One could speculate that the major reason for this difference is that Brener and Kleinman's subjects received a form of proportional feedback whereas Shapiro's subjects received binary feedback delivered both visually and aurally.

CARDIAC ARRHYTHMIAS

Although there have been several brief reports (Engel & Melmon, 1968; Weiss & Engel, 1970) on the operant conditioning of cardiac arrhythmias, most of the data seem to be summarized in one article (Weiss & Engel, 1971). The latter study contains detailed reports on eight patients suffering from premature ventricular contractions who were all treated in approximately the same experimental paradigm.

The paradigm was, for the most part, a replication of the procedures used in Engel's three previous studies. Patients were given binary visual feedback as to whether their heart rate was at or beyond the desired heart rate in the appropriate direction. In this study, contrary to the previous work with normals, patients were fully informed of all the details of the experiment, especially the response being monitored and also were not given monetary rewards for time accumulated in the correct direction. Despite the omission of this tangible reinforcer, the authors still describe their study in terms of "operant conditioning" rather than self-control through feedback.

Most patients were run under four conditions and some under a fifth. These were, in the usual order, heart rate speeding, heart rate slowing, alternately speeding and slowing heart rate for 1–4 minute trials, and then control of heart rate variability. In this final condition, subjects received binary feedback (light onset) when their heart rate stayed within a 10-beat range. The nature of this feedback arrangement also provided subjects with feedback of the occurrence of each premature ventricular contraction. The fifth condition, in which four of the patients were run, consisted of a gradual removal of feedback. Feedback was alternately presented and removed; the initial ratio of 1 minute of feedback to 1 minute without was extended to 1 minute on and 7 minutes off. Interestingly, in this final phase subjects did as well without feedback as with it, indicating that true self-control had been learned.

Four of the eight patients had a significantly reduced frequency of premature ventricular contractions after the training procedure as measured in the ward and at a follow-up 3–21 months afterward. A fifth patient did not have a significant reduction in premature ventricular contractions rate but did learn to recognize their occurrence and was able to control them at home through resting. The other three patients showed no evidence of learning to control premature ventricular contractions.

Total number of training sessions ranged 22–77 with a mean of 57.5. The five cases which can be considered successes all had 47 or more training sessions. Our reanalysis of these results show that improvement is significantly (by $p = .05$, Fisher's exact probability test) related to receiving 47 or more training sessions.

Weiss and Engel deserve commendation for the following two related techiques for extending their procedures into the area of clinical application: first, eliminating the necessity for external feedback to control the cardiac dysfunction, and more importantly, monitoring premature ventricular contraction rate outside the experimental chamber. The degree of generalization observed in this latter procedure is especially encouraging. Generalization of learned control of cardiac functioning to the natural environment, true self-control, is the desired clinical terminal behavior.

There are two obvious criticisms which can be leveled at this otherwise excellent clinical study. The first is the lack of a controlled group of patients who were either untreated or given some equally time-consuming and involving placebo task. This criticism is obviated to some degree because of the prolonged base-line period prior to the experimental phase during which premature ventricular contractions were noted. The documented history of premature ventricular contractions ranged 3 months–8 years with a mean of 3.0 years. Despite this, one would hope that a next step would be a controlled outcome study of the procedure.

The second criticism is the obvious confounding of several aspects of the training procedure. One does not know whether the entire sequence of training procedures is necessary or only part of it. Thus a next step would be to isolate which parts of the training procedure are necessary, facilitative, or superfluous, in order to improve efficiency.

ALTERNATIVE METHODS OF CARDIAC CONTROL

There are at least two additional lines of research which are related to the area of self-control of cardiac function: (*a*) the use of progressive relaxation as developed by Jacobson (1938) and (*b*) the use of various Yogic exercises and related forms of mediation. In both of these a significant degree of change in various cardiac functions has been produced.

Progressive Relaxation

Jacobson (1939) reported on the effects on blood pressure of training in and subsequent use of his system of muscle relaxation. His conclusions, which tended to be more qualitative than quantitative, were that there is a general relation between decreases in blood pressure and decreases in muscle activity as shown by the electromyogram and that training in "progressive relaxation" results in greater decreases in electromyogram activity than self-induced relaxation without training. He presented blood pressure data on 14 normotensive subjects, 9 of whom had been trained in progressive relaxation. These data, both systolic and diastolic blood pressure, were collected after a 15-minute adaptation period during which the subjects rested in a supine position and again after 45 minutes of self-induced relaxation. Analyses of these data by the present authors revealed significant decreases in both systolic blood pressure ($\Delta\bar{X} = 8.0$ millimeters of mercury; $F = 19.9$, $df = 1/12$, $p < .01$) and diastolic blood pressure ($\Delta\bar{X} = 7.8$ millimeters of mercury; $F = 11.5$, $df = 1/12$, $p < .01$), but no differences between the trained and untrained subjects in degree of reduction achieved. It should be noted that the magnitude of change in blood pressure in this study exceeds that reported by Shapiro and his colleagues in any of their studies of normotensive subjects (see Table 2) and compares favorably with the results reported by Brener and Kleinman (1970).

In a second study, Jacobson (1940) reported decreases in heart rate associated with relaxation training. Heart rate was measured initially after a 15-minute adaptation period during which the subjects rested in a supine position and later after 45 minutes of progressive relaxation. The average decrease in heart rate was 6.5 beats per minute. Analysis of this data *by the present authors* shows this change to be highly significant ($t = 5.59$, $p < .001$).

In a more recent study, Paul (1969) reported on physiological changes in college-age females who were treated for 30 minutes by either (*a*) a brief form of progressive relaxation training, (*b*) hypnosis induction and suggestions of relaxation, or (*c*) self-instructed relaxation. Heart rate data were collected during a 10-minute adaptation period and at the end of training. The progressive relaxation training group showed significantly greater changes in heart rate than the other two groups with reductions of approximately 8 beats per minute occurring in both sessions. It seems possible that Paul's subjects were not completely adapted when the initial heart rate recording was made which could account for part of the magnitude of change.

In a later report Paul and Trimble (1970) compared subjects given training in brief progressive relaxation either by tape recording or by a therapist present in the room. The latter condition led to significantly greater decreases in heart rate; however, the tape-recorded relaxation treatment did result in a group mean decrease in heart rate of approximately 4 beats per minute which is comparable to the results reported by most other investigators.

It thus seems that training in deep muscle relaxation, even of an abbreviated form, can lead to decreases in heart rate comparable to those achieved by most of the studies in the biofeedback area. This finding leads us to speculate that the operant conditioning or self-control procedures are possibly only inefficient methods of teaching subjects to relax.

Yoga

Although many claims of phenomenal degrees of self-control among Yogis had been made, few scientific data were collected on this subject until the 1950s when Wenger and his associates (Wenger & Bagchi, 1961; Wenger, Bagchi, & Anand, 1961) made systematic autonomic recordings from over 50 Yogis in India. In one study of three Yogis who claimed to be able to stop their heart, Wenger et al. (1961) found (*a*) absence of radial pulse, (*b*) absence of most heart sounds, (*c*) blood pressure radically increased, but (*d*) no stoppage of the heart and, in fact, (*e*) very little slowing in heart rate. They concluded that these subjects were utilizing thoracic muscle interference with venous return which led to a radical decrease in blood flow.

A fourth Yogi claimed only to be able to slow his heart rate and did so, demonstrating increases in interbeat intervals from a base line of approximately 1.0 second to slightly over 2 seconds for several beats (equivalent to changes in heart rate from 60 beats per minute to 30 beats per minute. The authors concluded that no muscle interference was apparently used; rather, the alterations seemed to be due to vagal innervation.

A similar example of voluntary cardiac arrest was reported by McClure (1959) in the case of a 44-year-old Caucasian male. This patient, who showed no cardiac irregularities, could voluntarily go into cardiac arrest for periods of approximately 5 seconds, increasing the interbeat interval of his heart rate from normally 1.0 second to over 5 seconds. Although the patient did not know the mechanism involved, McClure concluded that it was due to vagal innervation and was entirely unrelated to respiration rate.

Three systematic studies of the effects of Yoga exercises and meditation on cardiac function have recently been reported. Datey, Deshmukh, Dalvi, and Vinekar (1969) studied the effects of training in the Yogic exercise called "Shavasan" on 47 hypertensive patients. The patients, whose original blood pressures were in the range 160–270/90–145, all received training in Shavasan for unspecified lengths of time. The typical training course was described as requiring 3 weeks of daily 30-minute sessions. Results were reported for three groups of patients: (*a*) 10 not on medication; (*b*) 22 whose hypertension was controlled by drugs; and (*c*) 15 whose hypertension was not adequately controlled by drugs. In all cases blood pressure was reported as the diastolic pressure plus one-third of the pulse pressure. For those not on medication blood pressure was reduced in 9 out of 10 cases with the average reduction being 27 millimeters of mercury ($p < .05$). In the second group the effects of training in Shavasan were evaluated in terms of the reduction in patients' drug dosage necessary to keep their blood pressure under control. In 13 of 22 patients, the drug dosage was decreased by 33% or more, with an average reduction for those 13 patients of 68%. In the third group the effects were evaluated in the same fashion and drug dosage was reduced by at least 33% in 6 of 15 patients. This study seems to show that a chronic course of training in yogic exercise had a clinically significant effect on a majority of the hypertensive patients studied, and would warrant investigation in a controlled-outcome study.

In the first of two studies of similar design Wallace (1970) studied the physiological effects of "transcendental meditation," as taught by the Maharishi Mahesh Yogi. Physiological responses of 15 "normal" college students, who had been practicing "transcendental meditation" for from 6 months to 3 years, were monitored both for 30 minutes prior to a 20–30 minute meditation period and during the latter period. In the five subjects for whom heart rate was measured, an average decrease of 5 beats per minute was observed. It was not stated, however, whether or not the magnitude of change was statistically significant.

In a more elaborate study of similar design, Wallace, Benson, and Wilson (1971) monitored physiological responses of 36 subjects who had been practicing "transcendental meditation" for an average of 29 months. The principal results of comparisons between the meditation period and the 30 minutes immediately preceding, during which subjects were resting comfortably, were significant changes in electroencephalogram and oxygen consumption. There was also a small but significant ($p < .05$) decrease in heart rate ($\Delta\bar{X} = -.3$ beats per minute), but no change in blood pressure.

These studies, though not of the controlled-outcome format in which a "no-treatment" or "attention-placebo" group is monitored, are highly suggestive that Yoga and other Eastern forms of exercise-meditation training can have important clinical application in the field of cardiac self-control. Certainly the changes achieved are of sufficient magnitude to warrant exploring this line of research as well as the ones discussed previously.

CONCLUSIONS

The self-control of several cardiac functions, acceleration, deceleration, and variability of heart rate and increases and decreases in blood pressure, has been amply demonstrated in normal subjects. However, with few notable exceptions the magnitude of change obtained has been small, the duration for which control was manifested has been brief, and the generalization of the changes obtained in the experimental chamber rarely demonstrated. Future research in this area must follow two directions if the promise it holds out is to be fulfilled: (*a*) there is now a need for chronic studies involving many trials to obtain large-scale changes; (*b*) there is also a need for more direct application of this work to clinical populations. Finally, as was pointed out in the final section of this paper, alternative modes of obtaining changes in cardiac functioning such as training in progressive relaxation, transcendental meditation, or various Yoga exercises, may prove to be more fruitful approaches to obtaining clinically significant changes than the operant-conditioning–feedback techniques reviewed in the main body of this paper. Only comparative outcome studies, conducted on a chronic basis, can answer this question, but certainly it is deserving of an answer.

REFERENCES

Benson, H., Shapiro, D., Tursky, B., & Schwartz, G. E. Decreased systolic blood pressure through operant conditioning techniques in patients with essential hypertension. *Science,* 1971, **173**, 740–742.

Bergman, J. S., & Johnson, H. J. The effects of instructional set and autonomic perception on cardiac control. *Psychophysiology,* 1971, **8**, 180–190.

Bergman, J. S., & Johnson, H. J. Sources of information which affect training and raising of heart rate. *Psychophysiology,* 1972, **9**, 30–39.

Blanchard, E. B., & Young, L. D. Relative efficacy of visual and auditory feedback for self-control of heart rate. *Journal of General Psychology,* 1972, **87**, 195–202.

Blanchard, E. B., Young, L. D., & McLeod, P. G. Awareness of heart activity and self-control of heart rate. *Psychophysiology,* 1972, **9**, 63–68.

Brener, J. Heart rate as an avoidance response. *Psychological Record,* 1966, **16**, 329–336.

Brener, J., & Hothersall, D. Heart rate control under conditions of augmented sensory feedback. *Psychophysiology,* 1966, **3**, 23–28.

Brener, J., & Hothersall, D. Paced respiration and heart rate control. *Psychophysiology,* 1967, **4**, 1–6.

Brener, J., & Kleinman, R. A. Learned control of decreases in systolic blood pressure. *Nature,* 1970, **226**, 1063–1064.

Brener, J., Kleinman, R. A., & Goesling, W. J. The effects of different exposures to augmented sensory feedback on the control of heart rate. *Psychophysiology,* 1969, **5**, 510–516.

Church, R. M. Systematic effect of random error in the yoked control design. *Psychological Bulletin,* 1964, **62**, 122–131.

Datey, K. K., Deshmukh, S. N., Dalvi, C. P., & Vinekar, S. L. "Shavasan": A yogic exercise in the management of hypertension. *Angiology,* 1969, **20**, 325–333.

Engel, B. T., & Chism, R. A. Effect of increases and decreases in breathing rate on heart rate and finger pulse volume. *Psychophysiology,* 1967, **4**, 83–89. (a)

Engel, B. T., & Chism, R. A. Operant conditioning of heart rate speeding. *Psychophysiology,* 1967, **3**, 418–426. (b)

Engel, B. T., & Hansen, S. P. Operant conditioning of heart rate slowing. *Psychophysiology,* 1966, **3**, 176–187.

Engel, B. T., & Melmon, L. Operant conditioning of heart rate in patients with cardiac arrhythmias. *Conditional Reflex,* 1968, 3, 130.

Finley, W. W. The effect of feedback on the control of cardiac rate. *Journal of Psychology,* 1971, **77**, 43–54.

Frazier, T. W. Avoidance conditioning of heart rate in humans. *Psychophysiology,* 1966, **3**, 188–197.

Headrick, M. W., Feather, B. W., & Wells, D. T. Unidirectional and large magnitude heart rate changes with augmented sensory feedback. *Psychophysiology,* 1971, **8**, 132–142.

Hnatiow, M. Learned control of heart rate and blood pressure. *Perceptual and Motor Skills,* 1971, **33**, 219–226.

Hnatiow, M., & Lang, P. J. Learned stabilization of cardiac rate. *Psychophysiology,* 1965, **1**, 330–336.

Jacobson, E. *Progressive relaxation,* Chicago: University of Chicago Press, 1938.

Jacobson, E. Variation of blood pressure with skeletal muscle tension and relaxation. *Annals of Internal Medicine,* 1939, 12, 1194–1212.

Jacobson, E. Variation of blood pressure with skeletal muscle tension and relaxation. II. The heart beat. *Annals of Internal Medicine,* 1940, **13**, 1619–1625.

Johns, T. R. Heart rate control in humans under paced respiration and restricted movement: The effect of instructions and exteroceptive feedback. *Dissertation Abstracts International,* 1970, **30**, 5712–5713.

Lang, P. J., Sroufe, L. A., & Hastings, J. E. Effects of feedback and instructional set on the control of cardiac rate variability. *Journal of Experimental Psychology,* 1967, **75**, 425–431.

Levene, H. I., Engel, B. T., & Pearson, J. A. Differential operant conditioning of heart rate. *Psychosomatic Medicine,* 1968, **30**, 837–845.

McClure, C. M. Cardiac arrest through volition. *California Medicine,* 1959, **90**, 440.

Paul, G. L. Physiological effects of relaxation training and hypnotic suggestion. *Journal of Abnormal Psychology,* 1969, **74**, 425–437.

Paul, G. L., & Trimble, R. W. Recorded vs. "live" relaxation training and hypnotic suggestion: Comparative effectiveness for reducing physiological arousal and inhibiting stress response. *Behavior Therapy*, 1970, **1**, 285–302.

Schwartz, G. E. Voluntary control of human cardiovascular integration and differentiation through feedback and reward. *Science*, 1972, **175**, 90–93.

Schwartz, G. E., Shapiro, D., & Tursky, B. Learned control of cardiovascular integration in man through operant conditioning. *Psychosomatic Medicine*, 1971, **33**, 57–62.

Shapiro, D., Schwartz, G. E., & Tursky, B. Control of diastolic blood pressure in man by feedback and reinforcement. *Psychophysiology*, 1972, **9**, 296–304.

Shapiro, D., Tursky, B., Gershon, E., & Stern, M. Effects of feedback and reinforcement on the control of human systolic blood pressure. *Science*, 1969, **163**, 588–590.

Shapiro, D., Tursky, B., & Schwartz, G. E. Control of blood pressure in man by operant conditioning. *Circulation Research*, 1970, **26** (Suppl. 1:27–32.) (a)

Shapiro, D., Tursky, B., & Schwartz, G. E. Differentiation of heart rate and systolic blood pressure in man by operant conditioning. *Psychosomatic Medicine*, 1970, **32**, 417–423. (b)

Shearn, D. W. Operant conditioning of heart rate. *Science*, 1962, **137**, 530–531.

Sroufe, L. A. Learned stabilization of cardiac rate with respiration experimentally controlled. *Journal of Experimental Psychology*, 1969, **81**, 391–393.

Sroufe, L. A. Effects of depth and rate of breathing on heart rate and heart rate variability. *Psychophysiology*, 1971, **8**, 648–655.

Stephens, J. H., Harris, A. H., & Brady, J. V. Large magnitude heart rate changes in subjects instructed to change their heart rates and given exteroceptive feedback. *Psychophysiology*, 1972, **9**, 283–285.

Wallace, R. K. Physiological effects of transcendental meditation. *Science*, 1970, **167**, 1751–1754.

Wallace, R. K., Benson, H., & Wilson, A. F. A wakeful hypometabolic physiologic state. *American Journal of Physiology*, 1971, **221**, 795–799.

Weiss, T., & Engel, B. T. Operant conditioning of heart rate in patients with premature ventricular contractions. *Psychosomatic Medicine*, 1971, **33**, 301–321.

Wenger, M. A., & Bagchi, B. K. Studies of autonomic functions in practitioners of Yoga in India. *Behavioral Science*, 1961, **6**, 312–323.

Wenger, M. A., Bagchi, B. K., & Anand, B. K. Experiments in India on "voluntary" control of the heart and pulse. *Circulation*, 1961, **24**, 1319–1325.

(Received April 10, 1972)

Psychological Bulletin
1973, Vol. 79, No. 3, 164–167

COMMENTS ON RAWLINGS' NONORTHOGONAL ANALYSIS OF VARIANCE

JOHN E. OVERALL[1]
University of Texas Medical Branch Galveston

DOUGLAS K. SPIEGEL
University of South Dakota

This note is a reply to Rawlings' criticisms of a previous article by the present authors. The method of analysis proposed by Rawlings as correct is precisely equivalent to our method, which he criticized as incorrect. There are difficulties inherent in the computational procedures proposed by Rawlings for more complex designs, and programs for accomplishing the analyses are considered.

Rawlings (1972) recently criticized a method suggested by the present authors for the analysis of variance of data from nonorthogonal designs (Overall & Spiegel, 1969). Substantial misconceptions are apparent in his criticisms, since a numerical example shows that the method which he proposed as appropriate is precisely equivalent to our method that he criticized as incorrect. Our previous article compared three different methods used by psychologists for the analysis of data from nonorthogonal experimental designs, and it pointed out that two of the methods *do not* provide tests of main effects as main effects are usually conceived in analysis of variance. Rawlings' criticisms overlook this purpose and imply that we presented all three methods as appropriate generalizations of the conventional analysis of variance for orthogonal designs. This will surely mislead readers who fail to refer to our article. Our purpose in the previous article was to show that the alternative methods test different hypotheses.

After detailed comparisons, we previously concluded that our Method 2 is the appropriate generalization of conventional analysis of variance. It yields results identical to those provided by several statisticians as examples of nonorthogonal analysis of variance (Rao, 1965; Snedecor & Cochran, 1967). Nevertheless, Rawlings (1972) states that although the method does "provide tests of hypotheses, these tests are not tests of main effects [p. 373]." Ironically, he then recommends a method of patching together several different computer programs to obtain an analysis directly equivalent to our Method 2.

A numerical example demonstrates the equivalence of Rawlings' analysis and the analysis previously proposed by us as the most appropriate generalization of orthogonal analysis of variance. In Table 1 are data for a two-way analysis of variance with disproportionate cell frequencies. Our previous article presented results obtained from analyses of these same data by three different methods. The method which we concluded to be most directly analogous to the conventional analysis of variance was Method 2. Rawlings' proposals for nonorthogonal analysis of variance can be shown to be precisely equivalent to our Method 2, at least for the case that he chose.

First, Rawlings proposes to calculate the error sum of squares by summing the squared deviation of observed scores about cell means:

$$SS_E = \sum \sum (y_{ij} - \bar{y}_i)^2.$$

The means for the 12 cells in Table 1 were calculated, and the squared deviations of observed scores were then summed to yield the following:

$$SS_E = (61 - 62)^2 + (73 - 62)^2 + \cdots + (91 - 88.7)^2 = 2{,}269.9.$$

We previously obtained exactly this same value as a residual about the full regression model.

The BMD05V computer program (Dixon, 1968) was employed, as recommended by Rawlings, to obtain appropriate sums of squares for testing main effects. To obtain the appropriate main effect sums of squares, the interaction terms were *not* included in the design matrix. The sums of squares obtained

---

[1] Requests for reprints should be sent to John E. Overall, Department of Neurology and Psychiatry, University of Texas Medical Branch, Galveston, Texas 77550.

TABLE 1

DATA FOR TWO-WAY ANALYSIS OF VARIANCE

Effect	Effect		
	A_1	A_2	A_3
B_1	61 73 52	42 53	96 81 92
B_2	79 65 81	37 32 50	45 37
B_3	43 35	87 81 65	75 59
B_4	56 25 19 35	72 84	98 77 91

from the Dixon program were

$$SS_A = 2{,}729.6$$

and

$$SS_B = 779.0.$$

These sums of squares for main effects are the same as we previously obtained in Method 2 analysis of these same data.[2] The composite sum of squares due to regression on *both* main effects can also be obtained directly from the Dixon program by inclusion of an appropriate "hypothesis card."[3] It was found to be

$$SS_H = b'X'Y = 3{,}722.8.$$

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OBTAINED IN THREE STEPS FOLLOWING RAWLINGS' RECOMMENDED PROCEDURE

Source	SS	df	MS	F
A	2,729.6	2	1,364.8	12.0
B	779.0	3	259.7	2.3
AB	9,560.1	6	1,593.4	14.0
Error	2,269.9	20	113.5	
Total	15,552.9			

This latter sum of squares is Rawlings' $b'X'Y$ with $\gamma = 5$ degrees of freedom. It is required to compute the sum of squares for interaction in the manner proposed by Rawlings:

$$SS_{AB} = Y'Y - b'X'Y - SS_E$$
$$= SS_T - SS_H - SS_E.$$

Thus, in this example, the interactions sum of squares is

$$SS_{AB} = 15{,}552.9 - 3{,}722.8$$
$$- 2{,}270.0 = 9{,}560.1.$$

The complete analysis of variance that was pieced together by calculating the error sum of squares from deviations about cell means, calculating the sums of squares for main effects using BMD05V, and calculating the interaction sum of squares as a residual is shown in Table 2. These results are identical (within the margin of rounding errors) to those previously presented for analysis of these same data by Method 2 (Overall & Spiegel, 1969).[4]

It is of interest to examine some potential difficulties in the use of his recommendations for analyses involving more complex multiclassification designs. First, the calculation of an error term directly from deviations about cell means is not practical. In a 14-way factorial design, for example, many cells may be empty or may contain only a single observation, and thus deviations from the cell means would be undefined. It is frequently difficult even to define all of the possible cells in a complex multiclassification design. The correct sum of squares for error can be obtained as a residual in the complete regression analysis. It is correct even if the design contains empty cells, provided that reasonable assumptions can be made about the absence of certain higher order interactions.

---

[2] The reader should be aware that the sum of squares due to regression on all main effects simultaneously cannot be obtained by simply summing the separate main effect sums of squares in a nonorthogonal design. The composite sum of squares for regression on all main effects can be obtained as the difference between the total sum of squares and the residual from the main effects model, $SS_H = SS_T - SS_R$.

[3] Two typographical errors occur in our original article. The design matrix presented in Table 2 (p. 314) contains a transposition of 0 and 1 in the final two entries in the row containing score 37. This transposition was not present in our original matrix, and the analysis is correct as presented. The error terms for all three methods are identical; thus the extra R^2 in the error term for Method 1 (p. 316) is superfluous.

[4] Psychologists are familiar with the problem of interpreting main effects in the presence of significant interactions. In "fixed effect" factorial designs, one usually considers main effects only if interactions involving those factors are not significant (Snedecor & Cochran, 1967). This problem is no different in nonorthogonal analysis of variance.

While the conception advanced by Rawlings of a single simple interaction as representing "lack of fit" to the main effect model is perfectly acceptable, a very real problem arises in more complex multiclassification designs. The section of Scheffé (1959, pp. 112–116) cited by Rawlings is concerned specifically with a simple two-way design.

Suppose one were to undertake a 10-way factorial instead. There would be 45 simple interactions plus a larger number of higher order interactions. It would be hazardous to lump together all interaction sums of squares to test the composite deviation from the main effect model against the within-cells error, even though such a test may be valid under an appropriately specified model. The presence of as many as a dozen significant interactions could easily be masked by the 1,000-plus zero interactions with which they are combined. In complex analyses, it seems important to include in the model interaction terms that are hypothesized to be present, to break out the separate sums of squares, and to test the specific interaction effects separately. Interactions assumed a priori to be nonexistent can be pooled into the error term. Interactions that cannot be assumed a priori to be absent should be included in the model.

Implicit in Rawlings' proposed analysis is the testing of a composite interaction sum of squares which, if not significant, is pooled into the residual error for testing main effects. Due to the low power of the composite test against specific interaction effects, such procedure should not be recommended for complex multiclassification designs. The pros and cons of pooling have been discussed frequently with regard to orthogonal analysis of variance (Winer, 1962, p. 202), and the general problem is no different in the nonorthogonal case. The present writers do not favor pooling of interaction sums of squares merely because of failure to find them statistically significant, but this is perhaps only a conservative preference on our part. In any event, the question of pooling has no unique relevance in consideration of nonorthogonal analysis of variance.[5]

In concluding this comment, a brief consideration of computer programs may be in order. The BMD05V program recommended by Rawlings for calculating sums of squares for main effects is adequate for the entire analysis, provided one recognizes that it accomplishes a general linear model analysis that is precisely Method 1 of our previous article. As pointed out by Rawlings, Method 1 provides correct tests of interaction effects, and we have added that a correct sum of squares for error can also be obtained directly from the Method 1 analysis. If one wishes to accomplish a Method 2 analysis using BMD05V or any other Method 1 program, two passes are required. First, a complete model including main effects and interactions is specified. This yields sums of squares for interactions and for residual error. Next, a reduced model involving only main effects is specified. This yields sums of squares for main effects. If Rawlings were going to recommend use of BMD05V for a portion of the analysis, it is puzzling why he did not recommend its use for all of the analysis. Of course, if one desires a Method 2 analysis, which Rawlings accepts as *the* appropriate method, it would be nice to have a computer program which produces that analysis directly. Such a program has been provided by Overall and Klett (1972, pp. 458–468).

---

[5] In his accompanying reply to this note, Rawlings (1972) has modified his original proposal to include consideration of pooled error estimates and interpretation of main effects in the presence of significant interaction. These considerations were not included in his previous proposal of an analysis which we have demonstrated to be equivalent to our Method 2. He also grossly misrepresents our recommendations with respect to Method 2. We did not recommend Method 2 as most appropriate "when results suggest departure from additivity." We actually stated that "Method 2 is appropriate where one hopes to account for systematic variation in terms of simple additive main effects but wishes to test the significance of interactions as a safeguard against nonadditivity [p. 319]."

REFERENCES

Dixon, W. J. (BMD05V) In, *Biomedical computer programs.* Los Angeles: University of California Press, 1968.

Graybill, F. A. *An introduction to linear statistical models.* New York: McGraw-Hill, 1961.

Overall, J. E., & Klett, C. J. *Applied multivariate analysis.* New York: McGraw-Hill, 1972.

Overall, J. E., & Spiegel, D. K. Concerning least squares analysis of experimental data. *Psychological Bulletin*, 1969, **72**, 311–322.

Rao, C. R. *Linear statistical inference and its applications.* New York: Wiley, 1965.

Rawlings, R. R., Jr. Note on nonorthogonal analysis of variance. *Psychological Bulletin*, 1972, **77**, 373–374.

Scheffé, H. *The analysis of variance.* New York: Wiley, 1959.

Snedecor, G. W., & Cochran, W. G. *Statistical methods.* (6th ed.) Ames: Iowa State University Press, 1967.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Early publication received June 15, 1972)

Psychological Bulletin
1973, Vol. 79, No. 3, 168–169

COMMENTS ON THE OVERALL AND SPIEGEL PAPER

ROBERT R. RAWLINGS, JR.[1]

National Institute of Mental Health, Rockville, Maryland

This note is a reply to the article by Overall and Spiegel. This note demonstrates that the method proposed by Overall and Spiegel and the method proposed by Rawlings are not equivalent.

The purpose of this note is the clarification of certain erroneous interpretations concerning the nonorthogonal analysis of variance which appear in Overall and Spiegel (1973). In part, in this note I demonstrate that their Method 2 is not equivalent to the method described by Rawlings (1972) referred to here as Method 1′.

First, I discuss in terms of hypothesis testing, the three methods for performing the analysis of variance described by Overall and Spiegel (1969). The model used in the discussion is the two-factor interaction model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

in which all cells are filled. We denote the population cell means by μ_{ij} and the cell frequencies by n_{ij}, $i = 1, \ldots, I$, $j = 1, \ldots, J$. As is well known the test for an A effect in Method 1 is a test of the hypothesis that the $A_i = \bar{\mu}_{i.} - \bar{\mu}_{..}$ are equal for all i, or equivalently it tests the hypothesis that $\alpha_i^* = \alpha_i - \bar{\alpha}_. + \bar{\gamma}_{i.} - \bar{\gamma}_{..}$ are equal for all i. This hypothesis is not data dependent and it is based upon the traditional definition of main effects A_i.

From pages 304 and 310 of Searle (1971) it is observed that the Method 2 test for an A effect is a test of the hypothesis that for all i

$$\sum_{j=1}^{J} \left[n_{ij}\mu_{ij} - \sum_{i'=1}^{I} \frac{n_{ij}n_{i'j}}{n_{.j}} \mu_{i'j} \right] = 0$$

or equivalently a test of the hypothesis that for all i

$$\left(n_{i.} - \sum_{j=1}^{J} \frac{n_{ij}^2}{n_{.j}} \right) \alpha_i - \sum_{i' \neq i} \left(\sum_{j=1}^{J} \frac{n_{ij}n_{i'j}}{n_{.j}} \right) \alpha_{i'} + \sum_{j=1}^{J} \left(n_{ij} - \frac{n_{ij}^2}{n_{.j}} \right) \gamma_{ij} - \sum_{i' \neq i}^{I} \sum_{j=1}^{J} \left(\frac{n_{ij}n_{i'j}}{n_{.j}} \right) \gamma_{i'j} = 0.$$

From page 307 of Searle (1971) it is observed that the Method 3 test for an A effect is a test of the hypothesis that the

$$\frac{1}{n_{i.}} \sum_{j=1}^{J} n_{ij}\mu_{ij}$$

are equal for all i, or equivalently that the

$$\alpha_i + \frac{1}{n_{i.}} \sum_{j=1}^{J} n_{ij}(\beta_j + \gamma_{ij})$$

are equal for all i.

In Overall and Spiegel (1969) it is recommended that Method 2 is the most appropriate method of analysis when results suggest departure from the additive model. However, since the hypothesis corresponding to Method 2 is complex *and data dependent*, it is rather doubtful that it is useful. Also, in the Overall and Spiegel article it is suggested that Method 3 may be more appropriate in testing a hierarchy of effects. But since Method 3 has a hypothesis that is data dependent and dependent upon the B effect and interactions, it is doubtful that this method is very useful.

On pages 316–318 Searle (1971) discusses the difficulties associated with these hypotheses. When interactions are present, it appears to this author that Method 1 at least possesses a more meaningful interpretation as in Scheffé (1959) (the effects of the levels of one factor are averaged over the levels of the other factor).

I refer to the test for an A effect using only the noninteraction model as Method 1′. When there are no interactions it can be shown that Methods 1 and 2 are equivalent to the hypothesis that all α_i are equal, which is the same hypothesis tested by Method 1′. It can also be shown that when we have no interactions and proportionate cell frequencies, then Method 3 tests the same hypothesis. Since Methods 1, 2, and 1′ test the same hypothesis when there are no interactions, one should use that statistic which provides the maximum power for the test of the hypothesis. In the following, I discuss Method 1′ as an improvement of Method 1 for the test of main effects. Also, I describe the extension of Method 1′ to designs with more than two factors. In Rawlings (1972) it is demonstrated mathematically that for large N, Method 1′ will always have larger power than Method 1 in the nonorthogonal designs (the powers are equal in the orthogonal case). This fact was the basis for my suggestion to use Method 1′ rather than Method 1. Mattsson and Rawlings[2] investigated the effects of different nonorthogonal designs on the power of the above statistical tests, and the numerical experiments consistently support the above conclusion. By similar arguments one can conclude that for an analysis with more than two factors one should retain only the significant interactions in the model when testing for any "appropriate" main effects if one is to increase the power of the main effects tests.

---

[1] Requests for reprints should be sent to Robert R. Rawlings, Jr., Health Service sand Mental Health Administration, National Institute of Mental Health, 5600 Fishers Lane, Rockville, Maryland 20852.

If there are more than two factors, then Method 1′ is described as follows: (*a*) Test for interactions and main effects using the complete interaction model. For those factors possessing significant interactions with other factors the main effects tests can be interpreted as above. If interactions are unlikely then one can perform a lack-of-fit test instead. (*b*) Test the remaining factors for main effects using a model containing only those interactions which were found to be significant in *a* above. In order to maintain an experimentwise significance level α, one should split up α between *a* and *b*, say α_1 and α_2 such that $\alpha_1 + \alpha_2 = \alpha$ as in Miller (1966). In brief, Method 1′ is simply the use of Method 1 on two different interaction models.

If Method 2 is to be proposed as an improvement over Method 1 then the following problems remain to be resolved: (*a*) If there are significant interactions in the general N factor model, but there are no interactions involving Factor A, then show that the test of an A effect reduces to a test of the same hypothesis as that tested by Method 1. (*b*) Demonstrate mathematically that Method 2 has larger power than Method 1 for detection of main effects, at least under certain conditions.

REFERENCES

MILLER, R. G., JR. *Simultaneous statistical inference.* New York: McGraw-Hill, 1966.

OVERALL, J. E., & SPIEGEL, D. K. Concerning least squares analysis of experimental data. *Psychological Bulletin,* 1969, **72**, 311–322.

OVERALL, J. E., & SPIEGEL, D. K. Comments on Rawlings' nonorthogonal analysis of variance. *Psychological Bulletin,* 1973, **79**, 164–167.

RAWLINGS, R. R., JR. Note on nonorthogonal analysis of variance. *Psychological Bulletin,* 1972, **77**, 373–374.

SCHEFFÉ, H. *The analysis of variance.* New York: Wiley, 1959.

SEARLE, S. R. *Linear models.* New York: Wiley, 1971.

(Early publication received October 4, 1972)

---

[2] N. B. Mattson and R. R. Rawlings, Jr. *Problems of power in nonorthogonal models.* (Tech. Rep. CSB-SAS-2) Washington, D. C.: National Institute of Mental Health, 1972.

Psychological Bulletin
1973, Vol. 79, No. 3, 170–171

COMMENT ON JOE'S "COMMENT ON OVERALL AND SPIEGEL'S 'LEAST SQUARES ANALYSIS OF EXPERIMENTAL DATA' "

I. LEON SMITH[1]

University of Cincinnati

The general linear model is discussed with respect to the conditions of orthogonality and equal subjects per cell in an analysis of variance application.

Joe's (1971) discussion and attempted clarification of the Overall and Spiegel (1969) article contains a basic error. He suggested that in the special case of equal subjects per cell, $T'A'AT$ will be a superdiagonal matrix. However, his own illustration of a two-way design with two levels of A and three levels of B with equal rather than unequal subjects per cell does not support this statement. Consider his example with two subjects per cell:

$$\begin{array}{cc|c|c|c|} & & & B & \\ & & 1 & 2 & 3 \\ \hline & 1 & \begin{matrix} x_{111} \\ x_{211} \end{matrix} & \begin{matrix} x_{112} \\ x_{212} \end{matrix} & \begin{matrix} x_{113} \\ x_{213} \end{matrix} \\ \hline A & 2 & \begin{matrix} x_{121} \\ x_{221} \end{matrix} & \begin{matrix} x_{122} \\ x_{222} \end{matrix} & \begin{matrix} x_{123} \\ x_{223} \end{matrix} \\ \hline \end{array} .$$

In terms of the general linear model for this design,

$$X = \begin{pmatrix} x_{111} \\ x_{211} \\ x_{112} \\ x_{212} \\ x_{113} \\ x_{213} \\ x_{121} \\ x_{221} \\ x_{122} \\ x_{222} \\ x_{123} \\ x_{223} \end{pmatrix} \quad A = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \phi = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \alpha_1\beta_1 \\ \alpha_1\beta_2 \\ \alpha_1\beta_3 \\ \alpha_2\beta_1 \\ \alpha_2\beta_2 \\ \alpha_2\beta_3 \end{pmatrix} . \qquad [1]$$

Since he wishes to form the dummy variables, $(\alpha_1 - \alpha_2)$, $(\beta_1 - \beta_3)$, $(\beta_2 - \beta_3)$, $(\alpha_1 - \alpha_2) \times (\beta_1 - \beta_3)$, and $(\alpha_1 - \alpha_2)(\beta_2 - \beta_3)$ from A, the appropriate T matrix is:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} . \qquad [2]$$

The square symmetric $T'A'AT$ matrix for this example is:

$$\begin{bmatrix} 12 & 0 & 0 & 0 & 0 & 0 \\ & 12 & 0 & 0 & 0 & 0 \\ & & 8 & 4 & 0 & 0 \\ & & & 8 & 0 & 0 \\ & & & & 8 & 4 \\ & & & & & 8 \end{bmatrix} . \qquad [3]$$

The matrix, of course, is not diagonal, even in the case of equal subjects per cell.

---

[1] Requests for reprints should be sent to I. Leon Smith, who is now at Curriculum Research and Development Center in Mental Retardation, Ferkauf Graduate School of Humanities and Social Sciences, Yeshiva University, 55 Fifth Avenue, New York, New York 10003.

The reason for this rests in the failure to form orthogonal dummy variables or contrasts from A. In the example, AT equals:

$$\begin{array}{cccccc}
\mu & (\alpha_1-\alpha_2) & (\beta_1-\beta_3) & (\beta_2-\beta_3) & \begin{array}{c}(\alpha_1-\alpha_2)\\(\beta_1-\beta_3)\end{array} & \begin{array}{c}(\alpha_1-\alpha_2)\\(\beta_2-\beta_3)\end{array} \\
1 & 1 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 0 & 1 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 \\
1 & 1 & -1 & -1 & -1 & -1 \\
1 & -1 & 1 & 0 & -1 & 0 \\
1 & -1 & 1 & 0 & -1 & 0 \\
1 & -1 & 0 & 1 & 0 & -1 \\
1 & -1 & 0 & 1 & 0 & -1 \\
1 & -1 & -1 & -1 & 1 & 1 \\
1 & -1 & -1 & -1 & 1 & 1
\end{array} . \qquad [4]$$

In order for contrasts to be orthogonal, the columns of AT must be independent—the sum of products of elements of any one column times another must be zero. In the case of the two contrasts for the B effect, as well as the interactions, the product is 4, which is identical to the offdiagonal elements in Matrix 3. Thus, $T'A'AT$ will be a diagonal matrix when there are equal subjects per cell, but only when the contrasts or dummy variables are orthogonal. Since Joe selected nonorthogonal contrasts from A, his Equation 13 will not yield the appropriate diagonal elements of $T'A'AT$ when both of the above conditions are met.

REFERENCES

JOE, G. W. Comment on Overall and Spiegel's "least squares analysis of experimental data." *Psychological Bulletin*, 1971, **75**, 364–366.

OVERALL, J. E., & SPIEGEL, D. K. Concerning least squares analysis of experimental data. *Psychological Bulletin*, 1969, **72**, 311–322.

(Received May 24, 1971)

Psychological Bulletin
1973, Vol. 79, No. 3, 172–179

JUDGMENTS AND EXPLANATIONS AS CRITERIA FOR THE PRESENCE OF COGNITIVE STRUCTURES

CHARLES J. BRAINERD[1]

University of Alberta

The current methodological disagreement over whether subjects' judgments or their explanations on Piagetian tasks should be viewed as the minimum necessary evidence for the presence of underlying cognitive structures is reviewed and resolved. The principal features of Piaget's theoretical model of cognitive structure are summarized as a means of resolving the disagreement. Two conclusions are authorized by this summary: (*a*) Explanations constitute inappropriate bases for inferring the presence/absence of cognitive structure because they are subject to at least two sources of systematic Type II error. (*b*) Judgments constitute appropriate bases for inferring the presence/absence of cognitive structure, because—from the standpoint of Piagetian theory—they are not subject to any known source of systematic error.

In recent years, Piaget (e.g., 1970a, 1970b) has reemphasized a fact about his theory of intellectual development which has received notably insufficient attention in the Piaget replication literature. The theory is—in its most basic sense—an account of the ontogenesis of a particular theoretical construct, *cognitive structure*. By implication, all of the now-familiar Piagetian dependent variables (e.g., conservation, transitivity, class–inclusion), mental processes (e.g., inversion, reciprocity), and laws (e.g., equilibration) serve but two broad ends, to determine the *presence* and the *nature* of cognitive structures. It is with the former of these two matters that the present paper is concerned.

The number of replications of the widely scattered reports of Piaget's original studies of intellectual development grew steadily during the past decade (cf. Brainerd & Allen, 1971a; Flavell, 1963, 1970; Hooper, Goldman, Storck, & Burke, 1971, for reviews). Perhaps the most noteworthy feature of this literature is the series of objections, some trivial, but most justified, that have been raised regarding Piaget's "clinical method." From the time Braine's doctoral dissertation and a subsequent monograph based on it (Braine, 1959) appeared until the present, these objections have fallen into two categories, "stimulus" criticisms and "response" criticisms. The stimulus criticisms (e.g., Braine, 1959, 1962; Braine & Shanks, 1965a, 1965b; Hall & Kingsley, 1968; Smedslund, 1963, 1969) concern methods of task *presentation* and they include the usual sources of Type II error in studies of children's reasoning: inattentiveness, forgetting, misunderstanding instructions, etc. The response criticisms concern the specific behaviors which are used to infer the presence of cognitive structures, that is, how shall structures be known? To date, the principal point of contention on the response side has been whether children's *judgments* or *explanations* of these judgments are the minimum necessary evidence that the structure presumed to underlie the solution of some given task is in fact present.

The purposes of this paper are to provide both a brief, selective review and, more important, a resolution of the second source of controversy, the response criticisms. We shall see that reviews of both the salient features of Piaget's model of cognitive structure and currently available classificatory schemes for explanations permit a resolution in favor of judgment responses.

Scope of the Problem and Preliminary Definitions

Since this paper is selective in the matter of specific publications cited, some prefacing remarks are in order concerning the range of phenomena to which the judgment/explanation controversy applies and the importance of its resolution. Also, it is advisable to be clear at the outset about what is meant by a judgment and an explanation in the context of Piagetian tasks.

---

[1] Requests for reprints should be sent to Charles J. Brainerd, Department of Psychology, University of Alberta, Edmonton 7, Alberta, Canada.

Scope

Although the range of dependent variables to be considered is restricted, the methodological controversy which they illustrate arises in *all* studies involving tasks designed to assess Piagetian cognitive structures. Thus, all of the numerous tasks which Piaget has employed in his studies of the various levels of cognitive structuration are relevant. For present purposes, however, the range will be limited to the standard (Flavell, 1970, p. 1003) methods of determining the presence of the class and relational structures of the concrete-operational period: conservation problems, seriation problems, class-inclusion problems.

Regarding the need to settle this controversy, it is not, as some might suppose, necessary to resolve it for the sake of establishing precise age norms for the appearance of Piagetian cognitive structures. Although it was in connection with age norms that the controversy first appeared, age norms are trivial issues from the standpoint of Piagetian theory because the theory leaves wide latitude regarding them. Instead, it is the hypothesized *correspondences* and *sequences* between structures which are the crucial empirical predictions. The theory leaves no latitude in these matters. Given two cognitive structures A and B, either they are acquired in strict synchrony or one invariably precedes the other. It is from the perspective of a need for further research on correspondences and sequences in the acquisition of cognitive structures that the judgment/explanation is considered herein. We simply cannot hope to get on with the business of evaluating specific predictions until a consensus on response criteria is reached.

Definitions

Consider the standard Piagetian tasks for assessing the presence of concrete-operational structures: (*a*) conservation problems, (*b*) three-term seriation problems, (*c*) class-inclusion problems. In *a,* the child is shown two objects that are perceptually identical *and* equivalent with respect to some quantitative property (e.g., mass). One of the objects is altered perceptually but not quantitatively. The experimenter then determines whether or not the child understands that quantitative equivalence is conserved across the perceptual transformation. In *b,* the child is shown three objects (A, B, C) which differ on some quantitative dimension (e.g., length). The experimenter demonstrates the relationship between the members of two of the three possible pairs (e.g., $A < B$, $B < C$) and then determines whether or not the child can deduce the relationship between the members of the third pair. In *c,* the child is shown two or more numerically unequal sets of objects (e.g., six horses and three cows). The experimenter determines whether or not the child infers that the number of elements in the superordinate set (animals) must be greater than for either subordinate set.

Traditionally, inferences about the presence of the cognitive structures which the preceding tasks presumably assess have been based on one or both of two sorts of responses, judgments and explanations. A judgment is generated as follows: A dichotomous question is posed about some quantitative relation which holds between the objects under consideration. The question is dichotomous in that the domain of possible responses is limited to two alternatives—either the objects are the same or they are different. To illustrate, in the conservation problem outlined above the child would be asked to judge whether or not the perceptually dissimilar objects were quantitatively equivalent. Since judgment responses are a dichotomous random variable, many possible response modes, in addition to verbalization, can be employed (e.g., cf. Beilin, 1965, 1968). An explanation is generated as follows: Following dichotomous judgments, the experimenter poses questions of the form "Why?" "How do you know that?" Explanations obviously restrict the domain of possible response modes to one—verbalization.

ORIGINS OF THE CONTROVERSY AND BASIC ISSUES

In 1959, Braine published a monograph ostensibly devoted to replicating Piaget, Inhelder, and Szeminska's (1960) studies of two concepts, transitivity (Problem *b* above) and discrimination of ordinal position. Braine's measurement techniques were radically different from those reported by Piaget et al. In particular, extraneous verbal factors from both the stimulus side (by employing a reinforced pretraining procedure) and the response side (by defining satisfactory performance in terms of seeking hidden candy) were eliminated. Braine's innovations might have gone unchallenged if it had not been for the fact that his data were at odds with Piaget's in one important respect. While Piaget found that transitivity and ordinality were acquired between the ages of 7 and 8, Braine reported that they were acquired 2 years earlier. Braine concluded that such a large discrepancy seemed to undermine the credibility of Piagetian theory.

Braine's findings soon provoked a lively exchange with Smedslund on "transitivity criteria." Throughout the exchange, both authors focused largely on stimulus issues. Presentation methods were scrutinized meticulously: Smedslund (1963) enumerated three possible sources of Type I error and three possible sources of Type II error in "psychological diagnostics" and he argued that one of the sources of Type I error (solutions by irrelevant hypotheses) probably accounted for Braine's data. Smedslund presented replication data of his own that supported Piaget's norms. Braine (1964) replied that Smedslund's criticisms were "far fetched" in that although solutions by irrelevant hypotheses were possible, there was no evidence for their actual occurrence. Braine presented fresh data which failed to reveal such errors and which replicated his original findings. In a rejoinder, Smedslund (1965) advanced the additional charge of "pseudomeasurement." Employing pseudomeasurement as a point of departure, Smedslund (1966) reported further data in support of Piaget's norms. During the same interim, Braine applied his methods to conservation problems (Braine & Shanks, 1965a, 1965b) and found that conservation concepts also appeared 2 years earlier than Piaget had reported.

It remained for Gruen (1966) to observe that the discrepancy between Braine's and Smedslund's data might be due in part to the different response criteria that each employed: Braine's criterion for the presence of transitivity was judgments only, while Smedslund's (and Piaget's) criterion was judgments plus adequate explanations. Gruen proceeded to reanalyze the data of his doctoral dissertation (Gruen, 1965), which was concerned with number conservation, and he found, in line with his hypothesis, that more subjects were inferred to possess conservation when the judgment criterion was applied than when the explanation criterion was applied. On the basis of this finding, Gruen concluded that the judgment criterion is more consonant with non-Piagetian theories of cognitive development (e.g., Bruner, 1964), while the explanation criterion is more consonant with Piagetian theory. Subsequent research has either directly (Roodin & Gruen, 1970) or indirectly (Brainerd, 1971a, 1971b) supported the proposition that the judgment criterion is the less stringent of the two.

Thus, we come to the crux of the methodological issue which confronts us: What inferences are to be made about subjects who provide correct judgments on Piagetian tasks but not satisfactory explanations? By Braine's criterion, one either must infer that the relevant cognitive structure is present or commit Type II errors. By Smedslund's criterion, one either must infer that structure is absent or commit at least some Type I errors. In the two sections which follow, we shall review those points which can be teased out of Braine's and Smedslund's many "stimulus" arguments in support of their respective response criteria.

Braine's Position and Its Insufficiencies

Although the judgment criterion is associated with Braine's name herein, many other investigators also have employed this criterion almost exclusively (e.g., Bever, Mehler, & Epstein, 1968; Calhoun, 1971; Mehler &

Bever, 1967). Hence, the association is simply a matter of historical precedent.

As it turns out, all but one of the various methodological arguments which Braine has advanced are strictly peripheral to issue under consideration because they pertain only to stimulus matters. The argument which *is* relevant appeared first in Braine (1959):

> "The reason why Piaget's results differ from those reported here is not simply that there is a difference in experimental procedures, but is, the writer believes, because in designing his experiments Piaget fails to eliminate important variables *which are not involved in the definition of the processes he sets out to investigate* [p. 16, italics added]."

Later, Braine (1962) offered illustrations of the sort of response criteria which were thought to follow from this argument. Bruner, Goodnow, and Austin's (1956) concept formation "strategies" were cited as cases in point: Bruner et al. first recorded dichotomous judgments on a series of simple concept identification tasks and then searched the responses inductively to determine whether or not there were any reliable patterns. Patterns thus identified were viewed as underlying processes.

Braine's argument and illustration clearly constitute a less than sufficient rationale for employing a judgment criterion with *Piagetian* tasks. It is charged that Piaget's methods are inconsistent with the "definition of the processes he sets out to investigate," but no direct evidence that Braine's methods provide a better fit with the definition is offered. This yields a rather paradoxical situation: If one wishes to argue that Piaget's methods do not square with his theory, it would seem imperative that one provide both a scrupulously precise synopsis of the theory and an equally precise statement of how one's own methods are more clearly authorized. For some unknown reason, Braine has done neither of these things; perhaps this is one more example of Gordon Allport's old principle of functional autonomy of methodology in psychological writing. In any event, the failure both to state the theory of cognitive structures precisely and to deduce the judgment criterion from the theory makes the quoted argument appear to be little more than a standard plea for better operational definitions.

Smedslund's Position and Its Insufficiencies

The association of Smedslund's name with the explanation criterion also is largely a matter of historical precedent. Other advocates of this criterion, notably, some of Piaget's chief collaborators (Inhelder, Bovet, Sinclair, & Smock, 1966; Inhelder & Sinclair, 1969) have been even more vocal than Smedslund in their contention that "special attention should be paid to the child's justification of his answers [Inhelder & Sinclair, 1969, p. 5]." Nevertheless, it is Smedslund's methodological tracts which purport to supply the rationale for the criterion.

As noted above, Smedslund (1963, 1969) has enumerated three sources of Type I errors encountered with Piagetian tasks. These are guesses, direct perception, and irrelevant hypotheses. Since the first two can easily be eliminated by certain refinements in presentation techniques (cf. Smedslund, 1969), only the third source need concern us. Irrelevant hypotheses refers to the fact that, occasionally, subjects fix on irrelevant hunches which permit correct judgments but *not* correct explanations. In Problem *b* above, for example, the subject might label objects that are presented first "little." Subsequently, when asked which of the three objects is least, the subject judges that A is "because it came first." Hence, presentation order forms an irrelevant basis for correct judgment.[2]

Although irrelevant hypotheses have been invoked to justify the explanation criterion, they fall short of being an adequate rationale for two basic reasons. First, they—like the other sources of Type I error—can be resolved by stimulus refinements; studies by Murray and Youniss (1968) and by Youniss and Murray (1970) have eliminated such errors by, among other things, varying presentation orders.[3] Second, Smedslund's arguments are characterized by the same weakness as Braine's—insufficient attention to the theory of cognitive structures. Like Braine, Smedslund has failed to show how his criterion fits with the theory. Instead, he has advanced methodological arguments (Type I error) when the only adequate rationale is a theoretical one. Hence, even if the third source of Type I error could not be resolved via stimulus refinements, the criterion would still be arbitrary.

---

[2] Another argument which Smedslund occasionally has employed to justify the explanation criterion is that the subject must prove that he sees the *logical necessity* of a given Piagetian concept. The author does not deal directly with this argument in the present paper, because it simply does not apply to the entire range of Piagetian concepts—not even to the three to which this paper is restricted: Many Piagetian tasks involve only approximate physical properties of objects about which there is no logical necessity at all. This is especially true of conservation problems (cf. Piaget, 1961) whose "logic" is disproven in modern relativity theory. By implication, Smedslund's argument neither has been nor can be deduced from the theory—which acknowledges a difference between logical and physical experience.

THE THEORY OF COGNITIVE STRUCTURES AND A RESOLUTION OF THE CONTROVERSY

Cognitive Structures Defined

There are two sides to Piagetian structuralism (e.g., Piaget, 1970b). First, there is a metatheoretical definition of the concept of structure suitable for application to all traditional academic disciplines. Second, there is a theory of psychological structure. Only the latter will concern us here.

Piaget postulates two principal forms of psychological structure, perceptual structure and cognitive structure. Two distinctions are drawn between perceptual and cognitive structure. First, perceptual structures are said to undergo only *quantitative* change after birth, while cognitive structures are said to undergo *qualitative* change after birth. In short, perceptual development amounts to refinements in function, while cognitive development amounts to gross changes in architecture. Second, perceptual structures presumably evolve in the womb, while cognitive structures presuppose internal representation (Piaget, 1950, 1954) and, therefore, cannot begin to evolve until the second year of life.

Turning to the formal properties of cognitive structures, they are defined as internalized wholes whose laws of composition are mental operations and whose self-regulatory rule is the principle of equilibration. The *function* of a cognitive structure is to interpret (assimilate) reality so that the cognizer can behave intelligently. Different levels of assimilation specify different levels of structure; for example, the cognitive structures of the concrete-operational level (middle childhood) permit only the assimilation of concrete facts, while the cognitive structures of the formal-operational level (adolescence) permit the assimilation of purely conjectural information. Concerning the developmental *origins* of cognitive structures, Piaget has been interested in only two possibilities: Are the structures "logical"—in the sense of being engendered by all that is intelligent in human action—or are they linguistic—in the sense of being embodiments of the inherent structure of man's natural languages? In contrast with learning-oriented accounts of cognitive development (e.g., Kendler, 1963), Piaget opts for logic rather than language and maintains that cognitive structures originate in the consolidation of certain general properties of overt action (e.g., inversion, reciprocity). Piaget (1970b, Ch. 5) even has gone so far as to agree with Chomsky's (1966) claim that the structures of language itself are "logical." Flavell (1963) indirectly provided perhaps the most apt comment on this matter when he observed that "language behavior is here treated as a dependent variable with cognition as the independent variable [p. 271]."

---

[3] For the sake of completeness, it must be pointed out that the Murray-Youniss/Youniss-Murray studies do *not* prove the existence of irrelevant hypotheses in any strict sense. Instead, all that has been demonstrated is that an $A = B \neq C$ transitivity assessment paradigm is more difficult than the standard $A \neq B \neq C$ paradigm (Problem *b* above). While it is possible that this difference in difficulty stems from the susceptibility of the $A \neq B \neq C$ paradigm to solutions by irrelevant hypotheses, it is equally possible that this finding is caused by the fact that children find three-term series problems involving mixed symmetrical ($=$) and asymmetrical ($\neq$) relations more difficult than three-term series problems involving only symmetrical relations or only asymmetrical relations. (E.g., cf. C. J. Brainerd, Structures of thought in middle childhood: Recent research on Piaget's concrete-operational groupments. Paper read at the Third Annual Conference on Structural Learning, Philadelphia, March–April 1972.)

Resolution

Two conclusions, which entail a resolution of the judgment/explanation controversy follow from the preceding account of cognitive structures: (*a*) The theory not only fails to justify an explanation criterion, but it makes such a criterion seem highly inappropriate to the task of determining the presence of cognitive structures. (*b*) The judgment criterion seems well-suited to the task of determining the presence of cognitive structures. Conclusion *a* follows mainly from the postulated nonverbal (or, more generally, nonsymbolic) nature of Piaget's structures. If, to paraphrase Flavell (1963), language is strictly a "dependent variable" vis-à-vis cognitive structure, then lengthy verbal rationales of the judgments made on Piagetian tasks amount to *sufficient but not necessary* conditions for the presence of structure. More simply, it follows from the theory that the set of all subjects evidencing adequate explanations on a Piagetian task is a proper subset of the set of all subjects who possess the structure which the task assesses. Therefore, if one makes adequate explanations a necessary precondition for inferring the presence of structure, as Smedslund and some of the Genevans do, then one has built in a source of Type II error and virtually insured inflated age norms for the acquisition of specific structures.

Turning to Conclusion *b*, the suitability of the judgment criterion follows both from the assimilative function of cognitive structures and from their logical origins. Concerning the former, the basic proof for the presence of a given cognitive structure must, by definition, be a proof that an act of understanding appropriate to that level of structuration has taken place. By implication, if we wish to avoid being unduly conservative, then we require the *minimum necessary* evidence that an act of understanding has occurred. Obviously, both the judgment and explanation criteria presuppose such assimilative acts; however, the latter is more conservative than the former. Since neither criterion is subject to any known source of Type I error (by the present analysis), judgments are more satisfactory than explanations as a criterion for the presence of cognitive structures. Concerning the logical origins of cognitive structures, one should, in theory, opt for the response criterion which is the least restrictive regarding the domain of specific response systems. Again, the choice is clear: While judgments can incorporate many diverse response modes (e.g., verbalization, pointing, level pulling), explanations incorporate only the verbal mode of response. In short, if one chooses to employ the explanation criterion, then one unduly restricts the behavioral domain to which the theoretical construct (structure) applies. This, of course, amounts to another Type II error.

Before summing up, the author should like to offer two further reasons, one substantive, the other pragmatic, for preferring the judgment criterion over the explanation criterion. Speaking substantively, there is the matter of the very significant research on cognitive development in deaf subjects which Furth (e.g., 1964, 1971) has been conducting: Piaget long has maintained that one reasonable way to approach his hypothesis that cognitive structures originate in logic rather than language is via the study of deaf children. Piaget's major predictions are that both the specific structures *and* their order of emergence will be the same in deaf and normal children. Obviously, one must study judgment responses with the language-deficient deaf. Speaking pragmatically, there has been a notable lack of uniformity (to date) among the classification schemes which investigators have applied to explanation data stemming from Piagetian tasks. Unfortunately, little or no consensus exists among such investigators on the matter of what constitutes an adequate explanation. While there is extensive agreement that an adequate explanation must be preceded by a correct judgment, there is little agreement beyond this point. To illustrate this claim, consider two recent studies of quantity conservation (Problem *a* above), the first by Lloyd (1971) and the second by the writer (Brainerd, 1971b). Suppose that both studies included at least one subject who makes consistently correct conservation judgments (i.e., concludes that objects remain quantitatively invariant following perceptual deformation) and who explains his judgments as follows, "They are equal because you could always change the 'different' one back." By Lloyd's scheme, based on Greenfield's (1966) work, these explanations would be labeled "direct-action" and viewed as conceptually primitive. By Brainerd's scheme, derived from Piaget's concrete-operational theory, the same explanations would be denoted "inversion reversibility" and viewed as quite adequate. It is not necessary to pursue this matter further. Suffice it to say that until there is some consensual agreement on classification schemes, arguments for the use of the explanation criterion, no matter how trenchant—contain a large element of wishful thinking.

To summarize the support for the judgment criterion, although judgments run the usual risk of *extraneous* Type I and Type II errors which crop up in any measurement situation, one has not, from the standpoint of the theory, built in any known source of error. The same cannot be said for the explanation criterion and therein lies the difference between the two.

Conclusions

The author should like to advance two further conclusions. First, Braine's (1959) earlier claim that his procedural innovations fit the preconceptions of Piagetian theory appears to have been vindicated for the limited case of his response criterion. This conclusion admittedly is at variance with the existing literature on the judgment/explanation controversy. Gruen (1966) argued in his earlier review that the explanation criterion is more consistent with the presuppositions of Piagetian theory than the judgment criterion is. We have seen that this line of reasoning does not stand up under close scrutiny. Paradoxically, from the standpoint of Gruen's argument, it would seem that the explanation criterion is more in line with the language-oriented theories of Piaget's chief opponents (notably, Bruner) than with Piaget's own theory.

The second, more important, conclusion pertains to the question of what function is fulfilled by explanations if they are not to be used for inferring the absolute presence of cognitive structures. Although there surely are others, the author would like to suggest one obvious one from his own research (Brainerd, 1971a, 1971b; Brainerd & Allen, 1971b): Explanations can supplement judgments in such a way that one is provided with insights into the nature of the structure or structures under consideration. It should be recalled in this regard that, according to Piaget, language bears a very special relation to cognitive structure. In particular, language is a dependent variable vis-à-vis cognitive structure. Logically, this implies that *all* identifiable structural properties of language are, in some sense, inherent properties of the cognitive structures. This, in turn, implies that one can discover some, *though certainly not all,* of the properties of cognitive structures via the study of language. Obviously, those forms of language which appear in direct connection with the functioning of particular structures (i.e., explanations of judgments on Piagetian tasks) constitute potentially the most fruitful source of linguistic information about that structure. This final conclusion also is very much in line with the transformational grammarians' (e.g., Chomsky, 1966) approach to language itself and, therefore, it specifies an interface between Piagetian cognitive-developmental theory and modern developmental psycholinguistics.

REFERENCES

Beilin, H. Learning and operational convergence in logical thought development. *Journal of Experimental Child Psychology*, 1965, **2**, 317–339.

Beilin, H. Cognitive capacities of young children: A replication. *Science*, 1968, **162**, 920–921.

Bever, T. G., Mehler, J., & Epstein, J. What children do in spite of what they know. *Science*, 1968, **162**, 921–924.

Braine, M. D. S. The ontogeny of certain logical operations: Piaget's formulations examined by nonverbal methods. *Psychological Monographs*, 1959, **73**(5, Whole No. 475).

Braine, M. D. S. Piaget on reasoning: A methodological critique and alternative proposals. In W. Kessen & C. Kunlman (Eds.), Thought in the young child. *Monographs of the Society for Research in Child Development*, 1962, **27**(2, Whole No. 83).

Braine, M. D. S. Development of a grasp of transitivity of length: A reply to Smedslund. *Child Development*, 1964, **35**, 799–810.

BRAINE, M. D. S., & SHANKS, B. L. The conservation of a shape property and a proposal about the origin of the conservations. *Canadian Journal of Psychology*, 1965, **19**, 197–207. (a)

BRAINE, M. D. S., & SHANKS, B. L. The development of conservation of size. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 227–242. (b)

BRAINERD, C. J. The development of the proportionality scheme in children and adolescents. *Developmental Psychology*, 1971, **5**, 469–476. (a)

BRAINERD, C. J. Reinforcement and reversibility in quantity conservation acquisition. Unpublished manuscript, University of Alberta, 1971. (b)

BRAINERD, C. J., & ALLEN, T. W. Experimental inductions of the conservation of "first-order" quantitative invariants. *Psychological Bulletin*, 1971, **75**, 128–144. (a)

BRAINERD, C. J., & ALLEN, T. W. Training and generalization of density conservation: Effects of feedback and consecutive similar stimuli. *Child Development*, 1971, **42**, 693–704. (b)

BRUNER, J. S. The course of cognitive growth. *American Psychologist*, 1964, **19**, 1–15.

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking*. New York: Wiley, 1956.

CALHOUN, L. G. Number conservation in very young children: The effect of age and mode of responding. *Child Development*, 1971, **42**, 561–572.

CHOMSKY, N. *Cartesian linguistics*. New York: Harper, 1966.

FLAVELL, J. H. *The developmental psychology of Jean Piaget*. Princeton, N. J.: V. Nostrand, 1963.

FLAVELL, J. H. Concept development. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology*. New York: Wiley, 1970.

FURTH, H. G. Research with the deaf: Implications for language and cognition. *Psychological Bulletin*, 1964, **62**, 145–164.

FURTH, H. G. Linguistic deficiency and thinking: Research with deaf subjects 1964–1969. *Psychological Bulletin*, 1971, **76**, 58–72.

GREENFIELD, P. M. On culture and conservation. In J. S. Bruner, R. R. Olver, & P. M. Greenfield (Eds.), *Studies in cognitive growth*. New York: Wiley, 1966.

GRUEN, G. E. Experiences affecting the development of number conservation in children. *Child Development*, 1965, **36**, 963–979.

GRUEN, G. E. Note on conservation: Methodological and definitional considerations. *Child Development*, 1966, **37**, 977–983.

HALL, V. C., & KINGSLEY, R. Conservation and equilibration theory. *Journal of Genetic Psychology*, 1968, **113**, 195–213.

HOOPER, F. H., GOLDMAN, J. A., STORCK, P. A., & BURKE, A. M. Stage sequence and correspondence in Piagetian theory: A review of the middle-childhood period. *Research relating to children*. (Bulletin 28) Washington, D.C.: U.S. Government Printing Office, 1971.

INHELDER, B., BOVET, M., SINCLAIR, H., & SMOCK, C. D. On cognitive development. *American Psychologist*, 1966, **21**, 1–23.

INHELDER, B., & SINCLAIR, H. Learning cognitive structures. In P. H. Mussen, J. Langer, & M. Covington (Eds.), *Trends and issues in developmental psychology*. New York: Holt, 1969.

KENDLER, T. S. Development of mediating responses in children. In J. C. Wright & J. Kagan (Eds.), Basic cognitive processes in children. *Monographs of the Society for Research in Child Development*, 1963, **28**(2, Whole No. 86).

LLOYD, B. B. Studies of conservation with Yoruba children of differing ages and experience. *Child Development*, 1971, **42**, 415–428.

MEHLER, J., & BEVER, T. G. Cognitive capacity of very young children. *Science*, 1967, **158**, 141–142.

MURRAY, J. P., & YOUNISS, J. Achievement of inferential transitivity and its relation to serial ordering. *Child Development*, 1968, **39**, 1259–1268.

PIAGET, J. *The origins of intelligence in children*. New York: International Universities Press, 1950.

PIAGET, J. *The construction of reality in the child*. New York: Basic Books, 1954.

PIAGET, J. The genetic approach to the psychology of thought. *Journal of Educational Psychology*, 1961, 52, 275–281.

PIAGET, J. Piaget's theory. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology*. New York: Wiley, 1970. (a)

PIAGET, J. *Structuralism*. New York: Basic Books, 1970. (b)

PIAGET, J., INHELDER, B., & SZEMINSKA, A. *The child's conception of geometry*. New York: Basic Books, 1960.

ROODIN, M. L., & GRUEN, G. E. The role of memory in making transitive judgments. *Journal of Experimental Child Psychology*, 1970, **10**, 264–275.

SMEDSLUND, J. Development of concrete transitivity of length in children. *Child Development*, 1963, **34**, 389–405.

SMEDSLUND, J. The development of transitivity of length: A comment on Braine's reply. *Child Development*, 1965, **36**, 577–580.

SMEDSLUND, J. Performance on measurement and pseudomeasurement tasks by five- to seven-year-old children. *Scandinavian Journal of Psychology*, 1966, **7**, 81–92.

SMEDSLUND, J. Psychological diagnostics. *Psychological Bulletin*, 1969, **71**, 234–248.

YOUNISS, J., & MURRAY, J. P. Transitive inference with nontransitive solutions controlled. *Developmental Psychology*, 1970, **2**, 169–175.

(Received September 16, 1971)

Psychological Bulletin
1973, Vol. 79, No. 3, 180

COMMENT ON MISUSE OF COVARIANCE

D. A. SPROTT

University of Waterloo

In discussing my note on the analysis of covariance (Sprott, 1970), Harris, Bisbee, and Evans (1971) note that my correction of the "arithmetic" in Evans and Anastasio (1968) still left a "significant" F ratio of 12 (instead of their 1,789). It is curious, and perhaps symptomatic of the ritualistic way in which statistical methods are often interpreted, that a numerical error by a factor of 100 in an F ratio is regarded with equanimity. The force of the example is apparently unchanged; because it is still "significant," the example is regarded as fulfilling its original purpose. Equally curious is the claim that there is a "significant effect even when none was present." Significance is a property of the data; in the example it was present according to an objective numerical test. If no effect was *built into* the example, this merely demonstrates that the construction of the example was faulty (or that they were the victims of an event with probability .01). The purpose of my note was to show that there is no substitute for thought. Blanket condemnation can be as damaging as ritualistic acceptance. The analysis of covariance, even when treatment and covariate are correlated, can be useful and valid when in the hands of a scientist who knows his data, knows what he is doing, and is prepared to think about the interpretation of his analyses.

REFERENCES

Evans, S. H., & Anastasio, L. J. Misuse of analysis of covariance when treatment effect and covariate are confounded. *Psychological Bulletin,* 1968, **69,** 225–234.

Harris, D. R., Bisbee, C. T., & Evans, S. H. Further comments—Misuse of Covariance. *Psychological Bulletin,* 1971, **75,** 220–222.

Sprott, D. A. Note on Evans and Anastasio on the Analysis of Covariance. *Psychological Bulletin,* 1970, **73,** 303–306.

(Received September 23, 1971)

REPLY TO SPROTT

SELBY H. EVANS

Texas Christian University

The original purpose of the illustration (Evans & Anastasio, 1968) was to show that analysis of covariance could leave a statistically significant residual treatment effect even though there was no treatment effect in the variate not predictable from the covariate. An F ratio is not normally used as an index of anything (for an index of residual variance, one would ordinarily use a partial correlation coefficient) and its magnitude bears on decisions of statistical significance only in the test of whether the F ratio exceeds or fails to exceed the critical value. As to our statement that there is a "significant effect even when none was present," the context of the paragraph should make it clear that the word *none* refers to *"no treatment effect in the variate not predictable from the covariate."*

REFERENCE

Evans, S. H., & Anastasio, L. J. Misuse of analysis of covariance when treatment effect and covariate are confounded. *Psychological Bulletin,* 1968, **69,** 225–234.

(Received September 23, 1971)

Psychological Bulletin
1973, Vol. 79, No. 3, 181–184

A COMPARISON OF TWO PROCEDURES FOR TESTING MULTIPLE CONTRASTS

JANE PERLMUTTER AND JEROME L. MYERS[1]
University of Massachusetts

Scheffé's test for contrasts is frequently used when a relatively small set of contrasts is of interest. When these contrasts have been planned prior to the data analysis, the Bonferroni t statistic provides a reasonable alternative. Comparisons are made of critical values for the two tests.

Given the usual multicell experimental design, the general null hypothesis that treatment population means are equal is, at best, of minimal interest. The more basic question is: Which means, or subsets of means, differ from each other. Null hypotheses relevant to such questions are generally expressed as linear contrasts of the form:

$$\psi = \Sigma_j w_j \mu_j = 0. \qquad [1]$$

When the weights (w_j) are $1, -1, 0, 0, \ldots$, the contrasts are between two treatments (pairwise comparisons). Frequently, theory or data will suggest contrasts between two subsets of means. For example, the contrast might be between the mean of a pool of two treatment groups and the mean of a pool of three other groups. In such a case, the weights would be -3, -3, 2, 2, and 2 or any multiplicative transform of those weights by a constant.

As the number of contrasts tested increases, the probability of at least one Type I error, the experimentwise error rate (EW), also increases. There are a number of excellent discussions of the error rate problem, including reviews of procedures for controlling EW, (Games, 1971; Miller, 1966; Ryan, 1959). It is apparent from such reviews that a plethora of procedures has been proposed to aid the experimenter in his search for significant differences within sets of means. Of these, Scheffé's test (Scheffé, 1953) seems preferable (cf. Myers, 1972) assuming (*a*) that it is desirable to hold EW at or below some designated level (and we believe that it is), (*b*) that contrasts have been selected on a post hoc basis, and (*c*) complex contrasts (contrasts assigning nonzero w_j to more than two treatments) are of interest. Scheffé's test controls EW for the infinitely large set of possible contrasts, does not require equal group sizes, is relatively robust against nonnormality and heterogeneity of variance, and is based on the F statistic, thus using readily available tables. Where interest centers entirely on pairwise comparisons, the Newman-Keuls procedure provides an honest experimentwise error rate with more power than the Scheffé procedure.

Since the Scheffé test ensures a constant EW, independent of the number of significance tests carried out, it is an excellent tool for a post hoc search for significant contrasts. However, its strength—that it sets EW for the entire infinitely large set of possible contrasts—is also its weakness. The experimenter pays a price in power for the privilege of testing all possible contrasts at a designated EW level.

Dunn (1961) and Davis (1969) have suggested an alternative to the strategy of selecting contrasts post hoc and testing by the Scheffé procedure. They note that an investigator who specifies even a large number of contrasts in advance of the data collection and uses the Bonferroni t test for controlling EW for that designated finite set, may well have more power than if he had used the Scheffé test. The present investigators have replicated Dunn's work and extended it to the .10 level. The purpose of the present paper is to summarize the results of comparing the power of

---

[1] Requests for reprints should be sent to Jerome L. Myers, Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01002.

Tables giving the Bonferroni t critical factor for K equal to the maximum number of contrasts, the Scheffé critical factor, and the K where the critical factors are equal, as well as tables of the criterion t statistic for various values of K for EWs of .01, .05, and .1 can be obtained by addressing requests to Jerome L. Myers at the above address. Computations were supported by a grant of free computer time from the University Computer Center, University of Massachusetts, Amherst.

Bonferroni t and Scheffé methods and to note the availability of tables summarizing the power comparisons and providing criterion statistics for EWs of .01, .05, and .10. The extension of Dunn's work to the .10 level should prove useful if we accept Scheffé's (1959) recommendation that this level be used in testing contrasts. If such a recommendation appears heretical, consider that EW is conceptually very different from the usual α level for single tests and that the use of a level higher than the conventional α level provides increased power while still ensuring that individual tests are carried out with extremely low probabilities of Type I errors.

A simple index of the relative power of the two techniques is the ratio of their confidence interval widths. The Bonferroni t interval is dependent upon K, the number of contrasts, while the Scheffé interval is not. Confidence interval widths were computed for the Bonferroni t for maximum K (to be defined below) and compared with the Scheffé interval for the same error degree of freedom and EW. In addition, values of K were computed for which the intervals were equal in width.

TWO CONFIDENCE INTERVALS

The criterion statistic for the Scheffé test is $a-1$ (where a is the total number of groups) times the F required for significance with $a-1$ and error degrees of freedom, and probability less than or equal to EW. Thus, the Scheffé confidence interval width equals:

$$(2)\cdot(\sqrt{MS_{\text{error}}})\cdot(\sqrt{\Sigma_j w_j/n_j})\cdot(\sqrt{(a-1)F_{\text{EW},a-1,df_{\text{error}}}}). \quad [2]$$

The Bonferroni t method is used when the set of contrasts is planned prior to the collection of data. The EW is selected for the set of contrasts, and partitioned among them. That is, typically an alpha level of EW/K is used for each contrast; however, the experimenter has the option of partitioning the EW in any way he chooses, and thus can assign a greater portion of the EW and significantly greater power to the contrasts of most interest. This test will also usually be conservative, but as will be pointed out, less so than for the Scheffé procedure for a wide range of K values. The reason for this conservatism is that this method is based on the Bonferroni inequality which essentially states that:

$$\text{EW} \leq \Sigma_j \alpha_j. \quad [3]$$

When the contrasts are orthogonal the equality will hold approximately but when the contrasts are overlapping the test will be conservative. The criterion statistic is a two-tailed t test or a single degree of freedom F test, and the confidence interval width equals:

$$(2)\cdot(\sqrt{MS_{\text{error}}})\cdot(\sqrt{\Sigma_j w_j/n_j})\cdot(\sqrt{F_{\text{EW}/K,\,1,\,df_{\text{error}}}}). \quad [4]$$

For purposes of comparing the power of the two methods the terms common to both intervals can be dropped and the "critical factors" are thus $\sqrt{(a-1)F_{\text{EW},\,1,\,df_{\text{error}}}}$ and $\sqrt{F_{\text{EW}/K,\,1,\,df_{\text{error}}}}$ for the Scheffé and Bonferroni t tests, respectively. It is obvious that the Bonferroni t critical factor is dependent on K while the Scheffé critical factor is not.

Although an infinite number of sets of weights (and thus contrasts) can be associated with any group of means, a "maximum" number of contrasts was defined. This included comparisons between means of individual treatment groups or between means

TABLE 1

MAXIMUM NUMBERS OF CONTRASTS

Number of means	Maximum number of contrasts
3	6
4	25
5	90
6	301
7	966
8	3,025
9	9,330
10	28,501
11	86,526
12	261,625
13	788,970
14	2,375,101
15	7,141,686
16	21,457,825
17	64,439,010
18	193,448,101
19	580,606,446
20	1,742,343,625

Note. See text for description of how maximum number was arrived at.

of two sets of groups. Table 1 gives the maximum number of contrasts under this definition for from 3 to 20 groups. This number increases rapidly as the number of groups increases, but few experimenters would have theoretical reasons to be interested in nearly all of them. In fact, many investigators are only interested in pairwise comparisons $\left(\binom{a}{2}\right)$, which are considerably fewer.

To obtain critical factors for the two test statistics, a computer program was written to calculate the probability of any F given specified numerator and denominator degrees of freedom. Simpson's rule was used to evaluate the appropriate incomplete beta integral (Pearson, 1934) and a search routine was used to calculate the appropriate criterion F for any probability level within one-hundredth of the desired probability.

Comparison of the Bonferroni t critical factor for K equal to the maximum number of contrasts, as well as the Scheffé critical factor for EWs of .1, .05, and .01 indicates that the Scheffé value is always smaller than the Bonferroni t value for the maximum number of contrasts, and hence more powerful for this K value. However, as was pointed out, few experimenters would ever consider looking at so many contrasts, and thus it is of interest to see at what value of K the critical factors are equal.

The value of K where the Bonferroni t critical factor is equal to the Scheffé critical factor increases as a and error degrees of freedom increase, and as EW decreases, but is generally a rather large number (except perhaps for very small a where there are few possible contrasts). It therefore seems that power can be gained by planning even a liberally large number of contrasts of theoretical interest, and using the Bonferroni t method. Even a seemingly large number of planned contrasts is often fewer than the crossover K. However, even though the crossover K is often rather large, with substantial error degrees of freedom the interval widths for the two methods may be close enough to warrant buying the option of "data snooping" for a slight loss in power. Furthermore, if an experimenter plans a given number of contrasts, and uses the Bonferroni t method, but discovers after "data snooping" additional contrasts of interest, he may then use the Scheffé method to test the additional contrasts, with the consequence that his resulting EW is the sum of the EWs used for the two methods. Alternatively, he can abandon the Bonferroni t approach and use the Scheffé method for both the a priori and a posteriori contrasts, with a loss of power for the a priori contrasts.

Due to the recommendation to use the Bonferroni t method, and the lack of available tables for many significance levels, the following formula can be used to calculate the criterion t statistic using readily available two-tailed Z tables:

$$t_{\mathrm{EW}/K},\, df_{\mathrm{error}} = Z_{\mathrm{EW}/K} + \frac{Z^3_{\mathrm{EW}/K} + Z_{\mathrm{EW}/K}}{4\cdot(df_{\mathrm{error}} - 2)}. \quad [5]$$

The Newman-Keuls method for pairwise contrasts is also more powerful than the techniques which control error rate for the infinite set of possible contrasts (Scheffé & Tukey's honestly significant difference). This is so because successively less conservative tests are done for less divergent means. Unfortunately, no direct comparison can be made of the power of the Newman-Keuls and Bonferroni t methods, because the confidence interval becomes narrower for successive contrasts with the Newman-Keuls test.

In conclusion, it is suggested that the experimenter consider the power of multiple-comparison techniques prior to collecting data. It appears that the Bonferroni t or Newman-Keuls methods will often prove more powerful than other methods which control EW.

In addition, it is worth noting that the Bonferroni approach is not restricted to t and F tests. That is, the Bonferroni inequality $(\mathrm{EW} \leq \Sigma\alpha_j)$ also holds for other statistical tests such as chi-square and tests of correlation coefficients, and thus can be applied to control error rate in a variety of situations.

REFERENCES

Davis, D. J. Flexibility and power in comparisons among means. *Psychological Bulletin*, 1969, **71**, 441–444.

Dunn, O. J. Multiple comparisons among means. *Journal of the American Statistical Association*, 1961, **56**, 52–64.

Bonferroni *t* and Scheffé methods and to note the availability of tables summarizing the power comparisons and providing criterion statistics for EWs of .01, .05, and .10. The extension of Dunn's work to the .10 level should prove useful if we accept Scheffé's (1959) recommendation that this level be used in testing contrasts. If such a recommendation appears heretical, consider that EW is conceptually very different from the usual α level for single tests and that the use of a level higher than the conventional α level provides increased power while still ensuring that individual tests are carried out with extremely low probabilities of Type I errors.

A simple index of the relative power of the two techniques is the ratio of their confidence interval widths. The Bonferroni *t* interval is dependent upon K, the number of contrasts, while the Scheffé interval is not. Confidence interval widths were computed for the Bonferroni *t* for maximum K (to be defined below) and compared with the Scheffé interval for the same error degree of freedom and EW. In addition, values of K were computed for which the intervals were equal in width.

Two Confidence Intervals

The criterion statistic for the Scheffé test is $a-1$ (where a is the total number of groups) times the F required for significance with $a-1$ and error degrees of freedom, and probability less than or equal to EW. Thus, the Scheffé confidence interval width equals:

$$(2)\cdot(\sqrt{MS_{\text{error}}})\cdot(\sqrt{\Sigma_j w_j/n_j})\cdot(\sqrt{(a-1)F_{\text{EW},a-1,df_{\text{error}}}}). \quad [2]$$

The Bonferroni *t* method is used when the set of contrasts is planned prior to the collection of data. The EW is selected for the set of contrasts, and partitioned among them. That is, typically an alpha level of EW/K is used for each contrast; however, the experimenter has the option of partitioning the EW in any way he chooses, and thus can assign a greater portion of the EW and significantly greater power to the contrasts of most interest. This test will also usually be conservative, but as will be pointed out, less so than for the Scheffé procedure for a wide range of K values. The reason for this conservatism is that this method is based on the Bonferroni inequality which essentially states that:

$$\text{EW} \leq \Sigma_j \alpha_j. \quad [3]$$

When the contrasts are orthogonal the equality will hold approximately but when the contrasts are overlapping the test will be conservative. The criterion statistic is a two-tailed *t* test or a single degree of freedom F test, and the confidence interval width equals:

$$(2)\cdot(\sqrt{MS_{\text{error}}})\cdot(\sqrt{\Sigma_j w_j/n_j})\cdot(\sqrt{F_{\text{EW}/K},1,df_{\text{error}}}). \quad [4]$$

For purposes of comparing the power of the two methods the terms common to both intervals can be dropped and the "critical factors" are thus $\sqrt{(a-1)F_{\text{EW}},1,df_{\text{error}}}$ and $\sqrt{F_{\text{EW}/K},1,df_{\text{error}}}$ for the Scheffé and Bonferroni *t* tests, respectively. It is obvious that the Bonferroni *t* critical factor is dependent on K while the Scheffé critical factor is not.

Although an infinite number of sets of weights (and thus contrasts) can be associated with any group of means, a "maximum" number of contrasts was defined. This included comparisons between means of individual treatment groups or between means

TABLE 1

Maximum Numbers of Contrasts

Number of means	Maximum number of contrasts
3	6
4	25
5	90
6	301
7	966
8	3,025
9	9,330
10	28,501
11	86,526
12	261,625
13	788,970
14	2,375,101
15	7,141,686
16	21,457,825
17	64,439,010
18	193,448,101
19	580,606,446
20	1,742,343,625

Note. See text for description of how maximum number was arrived at.

of two sets of groups. Table 1 gives the maximum number of contrasts under this definition for from 3 to 20 groups. This number increases rapidly as the number of groups increases, but few experimenters would have theoretical reasons to be interested in nearly all of them. In fact, many investigators are only interested in pairwise comparisons $\left(\binom{a}{2}\right)$, which are considerably fewer.

To obtain critical factors for the two test statistics, a computer program was written to calculate the probability of any F given specified numerator and denominator degrees of freedom. Simpson's rule was used to evaluate the appropriate incomplete beta integral (Pearson, 1934) and a search routine was used to calculate the appropriate criterion F for any probability level within one-hundredth of the desired probability.

Comparison of the Bonferroni t critical factor for K equal to the maximum number of contrasts, as well as the Scheffé critical factor for EWs of .1, .05, and .01 indicates that the Scheffé value is always smaller than the Bonferroni t value for the maximum number of contrasts, and hence more powerful for this K value. However, as was pointed out, few experimenters would ever consider looking at so many contrasts, and thus it is of interest to see at what value of K the critical factors are equal.

The value of K where the Bonferroni t critical factor is equal to the Scheffé critical factor increases as a and error degrees of freedom increase, and as EW decreases, but is generally a rather large number (except perhaps for very small a where there are few possible contrasts). It therefore seems that power can be gained by planning even a liberally large number of contrasts of theoretical interest, and using the Bonferroni t method. Even a seemingly large number of planned contrasts is often fewer than the crossover K. However, even though the crossover K is often rather large, with substantial error degrees of freedom the interval widths for the two methods may be close enough to warrant buying the option of "data snooping" for a slight loss in power. Furthermore, if an experimenter plans a given number of contrasts, and uses the Bonferroni t method, but discovers after "data snooping" additional contrasts of interest, he may then use the Scheffé method to test the additional contrasts, with the consequence that his resulting EW is the sum of the EWs used for the two methods. Alternatively, he can abandon the Bonferroni t approach and use the Scheffé method for both the a priori and a posteriori contrasts, with a loss of power for the a priori contrasts.

Due to the recommendation to use the Bonferroni t method, and the lack of available tables for many significance levels, the following formula can be used to calculate the criterion t statistic using readily available two-tailed Z tables:

$$t_{\mathrm{EW}/K,\, df_{\mathrm{error}}} = Z_{\mathrm{EW}/K} + \frac{Z^3_{\mathrm{EW}/K} + Z_{\mathrm{EW}/K}}{4\cdot(df_{\mathrm{error}} - 2)}. \quad [5]$$

The Newman-Keuls method for pairwise contrasts is also more powerful than the techniques which control error rate for the infinite set of possible contrasts (Scheffé & Tukey's honestly significant difference). This is so because successively less conservative tests are done for less divergent means. Unfortunately, no direct comparison can be made of the power of the Newman-Keuls and Bonferroni t methods, because the confidence interval becomes narrower for successive contrasts with the Newman-Keuls test.

In conclusion, it is suggested that the experimenter consider the power of multiple-comparison techniques prior to collecting data. It appears that the Bonferroni t or Newman-Keuls methods will often prove more powerful than other methods which control EW.

In addition, it is worth noting that the Bonferroni approach is not restricted to t and F tests. That is, the Bonferroni inequality ($\mathrm{EW} \leq \Sigma\alpha_j$) also holds for other statistical tests such as chi-square and tests of correlation coefficients, and thus can be applied to control error rate in a variety of situations.

REFERENCES

DAVIS, D. J. Flexibility and power in comparisons among means. *Psychological Bulletin*, 1969, **71**, 441-444.

DUNN, O. J. Multiple comparisons among means. *Journal of the American Statistical Association*, 1961, **56**, 52-64.

GAMES, P. A. Multiple comparisons of means. *American Educational Research Journal*, 1971, 8, 531–565.

MILLER, R. G., JR. *Simultaneous statistical inference*. New York: McGraw-Hill, 1966.

MYERS, J. L. *Fundamentals of experimental design*. (2nd ed.) Boston: Allyn & Bacon, 1972.

PEARSON, K. *Tables of the incomplete B-function*. Cambridge: Cambridge University Press for the Biometrika Trustees, 1934.

RYAN, T. A. Multiple comparisons in psychological research. *Psychological Bulletin*, 1959, **56**, 26–47.

SCHEFFÉ, H. A. A method of judging all contrasts in the analysis of variance. *Biometrika*, 1953, **40**, 87–104.

SCHEFFÉ, H. A. *The analysis of variance*. New York: Wiley, 1959.

(Received October 12, 1971)

Manuscripts Accepted for Publication in the

Psychological Bulletin

Birth Order Effects: A Reply to Schooler. Hunter M. Breland (Educational Testing Service, Princeton, New Jersey 08540).

Learning, Transfer, and Changes in Ability Factors: A Multivariate Model. Allan R. Buss (Center for Advanced Study in Theoretical Psychology, University of Alberta, Edmonton, Alberta, Canada T6G 2E9).

The Random-Sampling Assumption in "Comment on Component-Randomization Tests. Eugene S. Edington (Department of Psychology, University of Calgary, Calgary, Alberta, Canada).

Criterion Models and Construct Validity for Criteria. Lawrence R. James (Department of the Navy, Navy Medical Neuropsychiatric Research Unit, Operational Psychiatry Division, San Diego, California 92152).

A Note on Minority Group Test Bias Studies. Richard R. Reilly (Educational Testing Service, Princeton, New Jersey 08540).

Psychological Bulletin
1973, Vol. 79, No. 3, 185–199

SELF- AND INTERPERSONAL EVALUATIONS:

ESTEEM THEORIES VERSUS CONSISTENCY THEORIES

STEPHEN C. JONES[1]

Cornell University

Two social psychological theories are contrasted and different predictions are derived concerning the relation between self and interpersonal evaluations. The prediction from self-esteem theory is that the higher an individual's evaluation of himself, the *less* his tendency to reciprocate evaluations from others, whereas the prediction from self-consistency theory is that the higher the individual's evaluation of himself, the *greater* his tendency to reciprocate evaluations from others. Studies bearing on these predictions are reviewed, and the evidence tends to support self-esteem theory. Furthermore, two extensions of the assumptions of self-esteem theory are presented and discussed in terms of accounting for results which apparently support self-consistency theory. Finally, whether or not the experimental subject is the direct target of another person's actions appears to distinguish esteem results from consistency results. The implications of this distinction for other problems of person perception and social evaluation are discussed. A general conclusion is that cognitive consistency theories may be somewhat overworked as explanatory frameworks for the study of social evaluations.

Although the interpersonal events involving the evaluations an individual receives from other people occur experimentally and in the "real world" with some regularity, confusions arise in attempting to explain these regularities. We observe that people typically talk more (e.g., Bavelas, Hastorf, Gross, & Kite, 1965; Jones, 1968b) appraise themselves more highly (e.g., Videbeck, 1960), and like others better (e.g., Aronson & Worchell, 1966; Backman & Secord, 1959; Jones & Panitch, 1971) when they receive approval from others than when they receive disapproval, but we know surprisingly little about why these things happen. There are wide differences of opinion about what underlying motivational processes affect a person's reactions to the knowledge that he is liked or disliked, that his opinions are valued or worthless, that he is accepted as a co-worker or rejected as a roommate, or in other ways praised or derogated by others.

The bulk of the investigations relating to this motivational issue unfortunately have been couched in terms of "mini theories" (Aronson, 1969), and attempts at more integrated approaches which go beyond a few intuitively appealing variables, or attempts to force a confrontation among various approaches have been all too few and far between. The purpose of this paper is to focus attention on two broad traditions in social psychology which relate to problems of interpersonal evaluations—self-esteem theories and cognitive consistency theories. Using as a focal point the question of how an individual reacts to evaluations received from others as a function of his self-evaluation, this paper discusses the distinctions between these theoretical positions, evaluates the evidence bearing on these distinctions, and suggests possible directions for resolving the differences.

THEORIES

"I am good/You love me/therefore you are good. . . . /I am bad/You love me/therefore you are bad." Does Laing's [1970, p. 10] poem of a schizophrenic relation represent a typical, self-consistent pattern of interpersonal evaluations? Or, if I think I am bad and discover that you love me, am I pleased as punch to get a little affection along the way? This "knot" of Laing's strikes at the heart of the controversy between esteem

---

[1] Requests for reprints should be sent to Stephen C. Jones, Department of Psychology, Cornell University, Ithaca, New York 14850.

TABLE 1

PREDICTED FAVORABILITY OF RATINGS OF OTHERS AS A FUNCTION OF SELF-EVALUATION AND APPRAISALS RECEIVED FROM OTHERS

Self-evaluation	Appraisals from others			
	Self-consistency theory		Self-esteem theory	
	Positive	Negative	Positive	Negative
High	+	—	+	—
Low	—	+	++	— —

and consistency approaches to the relation between self and interpersonal evaluations.

Self-Consistency Theories

The central notion of self-consistency theorists is that an individual's actions, attitudes, and his receptivity to information from other people are strongly affected by a tendency to create and maintain a consistent cognitive state with respect to his evaluations of himself. Various reasons are given for this cognitive tendency including economy in the organization of one's perceptions (Heider, 1958), the reduction of dissonance (Festinger, 1957), predictability in relationships with others (Newcomb, 1961; Secord & Backman, 1965), or avoiding cognitions with conflicting implications for action (Jones & Gerard, 1967).

As elaborated by Secord and Backman (1961, 1964, 1965), a state of self-consistency or congruency is said to exist "when [his own and others'] behaviors imply definitions of self congruent with relevant aspects of the self-concept [1961, p. 23]." Given a state of inconsistency or incongruency involving relationships with another person, the individual may change his conception of himself, change or misperceive his own actions, or in a variety of ways transform his relationship with the other person. Concerning the latter, he may interact only with those whose behavior validates his self-concept or misperceive an attempt to change those actions of the other person which produce the inconsistency. Finally, Secord and Backman (1961) propose that inconsistencies may be eliminated by selective evaluation of other people, and they predict that the individual "tends to increase his liking for [others] who behave toward him in a congruent fashion, and to decrease his liking for those who behave in an incongruent manner [p. 25]."

The self-consistency notion has played a role in other theoretical approaches to interpersonal relationships such as Heider's (1958) balance theory, Newcomb's (1961) symmetry model, and Baron's (1966) social reinforcement theory. It also has been central to the thinking of humanistically oriented self-concept theorists such as Lecky (1945) and Rogers (1951, 1959). Although these positions contain many differences in their foci and in their implications for action involving self and others, the common characteristic which concerns us is the proposal that relations between evaluations of the self and others are mediated by a tendency toward self-consistency.

To clarify this point and anticipate its distinction from the self-esteem theories, consider an experimental paradigm in which subjects have either high self-evaluations or low self-evaluations and receive either positive evaluations or negative evaluations from other members of their group. The question is to what extent the subjects accept or reject their peers. For individuals with high evaluations of themselves or some aspect of self, positive evaluations from others are consistent and negative evaluations are inconsistent, whereas for individuals with low self-evaluations, positive evaluations are inconsistent and negative evaluations are consistent. Therefore, the prediction from the self-consistency theories is that high self-evaluators will react more favorably to approval than to disapproval and that low self-evaluators will react more favorably to disapproval than to approval. This prediction is described in Table 1.

Self-Esteem Theories

The self-esteem position assumes that the individual has a need to enhance his self-evaluation and to increase, maintain, or confirm his feelings of personal satisfaction, worth, and effectiveness. Although this need is assumed to be general, at any given point in time it may manifest itself with respect to a particular aspect of one's self-evaluation rather than to more global feelings about the self. Furthermore, the state of the need varies with the degree of personal satisfaction or frustration the individual experiences in a particular situation or period of time. The self-esteem need also varies across individuals. It is assumed that this individual difference variation is reflected in attitudinal measures of self-esteem and that persons with high self-esteem are relatively more satisfied with respect to this need than persons with low self-esteem.

As was the case with the self-consistency theories, the self-esteem need is responsive to evaluative information the individual gains from his own behavior and comparative or reflected appraisals from other people. Coping effectively with the tasks and problems one encounters in his physical and social environment as well as gaining information from others that he is liked and respected or that his actions or characteristics are highly evaluated produce satisfactions for his self-esteem need. Data from a variety of social psychological studies, including, for example, studies of social influence (Hovland & Janis, 1959), self-presentation (Jones, 1964) and social comparison (Morse & Gergen, 1970), have been interpreted in terms of this direct relation between self-esteem needs and social acceptance or rejection.

The critical question for purposes of comparison with self-consistency theories is what predictions self-esteem theories make regarding the individual's reactions to positive or negative evaluations received from other people. In concert with the self-consistency theories, the major sources of change in thought or action involve the individual's evaluation of himself, his choice of activities or social roles, and his relationships with others. The difference is that these changes are designed to enhance self-esteem rather than to achieve self-consistency.

In general, one would expect people to respond favorably (i.e., like, benefit, agree with, etc.) to positive evaluations of themselves which are assumed to satisfy esteem needs and to respond unfavorably (i.e., dislike, aggress against, disagree with, etc.) to negative evaluations of themselves which are assumed to frustrate esteem needs. This prediction follows from the social exchange notions of Homans (1961) and Thibaut and Kelley (1959), the notion that attraction to a group is directly proportional to the needs satisfied by the group (Cartwright & Zander, 1960; Dittes, 1959), as well as the frustration–aggression hypothesis (Dollard, Doob, Miller, Mowrer, & Sears, 1939). Furthermore, this tendency should interact with the magnitude of the esteem needs as experimentally manipulated or measured. Specifically, low self-esteem individuals are predicted to respond more favorably to positive evaluations from others and more unfavorably to negative evaluations from others as compared to high self-esteem individuals. This prediction, first articulated by Dittes (1959), follows from the assumption that low self-esteem people have greater needs for esteem enhancement and are therefore more satisfied by the approval of others and more frustrated by the disapproval of others than are high self-esteem people. The prediction is described in Table 1.

Empirical Evidence

The critical test between the self-esteem and self-consistency theories involves manipulating or measuring the individual's evaluation of himself and the evaluations he receives from other people, then measuring the feelings he expresses to or about his associates, their characteristics or actions. At the present time there are a substantial number of studies which meet in whole or in part the conditions necessary for this critical test.

Self-Consistency Findings

Consider, first, the research which generally is viewed as supporting the self-consistency position. Three of the six studies are correlational in design, and they differ from other investigations described in that all the evaluative components measured are reports made by the subject. In particular, the evaluations purportedly received by the subject from others are *estimates* as to how he thinks others are evaluating him.

In a study of fourth-, fifth-, and sixth-grade school children of both sexes, Wiest

(1965) measured global self-esteem and correlated these measures with indices of the degree of congruency existing between a student's attraction for each other member of his group and his perception of others' feeling toward himself. Wiest obtained a significant positive correlation (.23) which he had predicted from Heider's balance theory. As self-esteem increased, interpersonal congruency also increased. Backman and Secord (1962) asked each member of a college sorority to select from a list of 16 adjective pairs the 5 adjectives most characteristic of herself, to indicate which 5 adjectives each other sorority member would attribute to her, and to rank order the other members according to their attractiveness. Interpersonal congruency was measured by the degree of overlap between the adjectives a subject selected for herself and those she estimated would be attributed to herself by others. In support of the Secord and Backman (1961) model, more congruency was obtained for the subject's ratings of the sorority members most liked by the subject than for the members least liked by the subject. From results obtained in a longer range investigation, Newcomb (1956) reported that interpersonal attraction was closely related to the perceived agreement between a person's self-description and his estimate of how other members of the dormitory would describe him.

In experimental studies as contrasted with correlational studies of the self-consistency theory, evaluations from others (which were manipulated) actually were received rather than estimated by the subject. In one of the three studies which partially fulfill the design conditions described above, Wilson (1965) led subjects to believe that they had failed on a task. Then either by having subjects make their own decisions (personal decision condition) or by having these decisions made for them by a "chance" procedure (chance decision condition), he brought about the decision that subjects would not proceed with a second task. After this decision, subjects received a bogus comment supposedly written by their partners, which in half of the conditions described a subject as competent and implied that he should accept the second task, and in half of the conditions described him as incompetent and implied that he should not accept the task. Wilson found that of the subjects in the personal decision conditions, those receiving a negative note rated their partners as more accurate and slightly more attractive than those receiving a positive note. For subjects in the chance decision condition, those receiving a positive note rated their partners as more emotionally supportive, more attractive, and slightly more accurate than those receiving the negative note.

Although this study is often cited as support for the self-consistency position (e.g., Aronson & Mettee, 1968; Secord & Backman, 1965), there is the question of why chance decision subjects who had also failed the first task and presumably held negative evaluations of their own abilities actually preferred partners who evaluated them positively but inconsistently. What most clearly distinguishes the chance from personal decision conditions is not a difference in level of self-evaluation but whether or not the subject made a *personal* decision not to continue. Therefore, although self-evaluative processes may be involved in the study, the results can be interpreted quite easily by assuming that individuals tend to like better and view as more accurate others who agree with their decisions as compared with those who disagree (cf. Festinger, 1954, 1957).

The most widely cited study of self-consistency and interpersonal evaluations is Deutsch and Solomon's (1959) test of derivations from Heider's (1958) balance theory. In their 2 × 2 design, self-evaluation was manipulated by informing subjects that they had either succeeded or failed on each of two tasks, and evaluations from others were manipulated by having each subject receive a note from one of her teammates stating either (e.g.) "You are the person I most prefer to have on my team again" or (e.g.) "You are the person I least prefer to have on the same team with me again." Analyses of evaluative ratings each subject made of her note-sending teammate showed, first, a "positivity" effect, that is, a tendency for subjects to rate more favorably the positive note sender than the negative note sender, and second, a self-consistency effect, that is, an interaction between self-evaluation and evaluations received from others in which the tendency for such a positivity effect was obtained only for subjects in the high-self-evaluation treatment. Low-self-evaluators tended to rate the negative note sender slightly, though not significantly, more favorably.

This experiment has not gone uncriticized. Pepitone (1966) reinterpreted their data by arguing that subjects may have been attempting to fulfill the announced purpose of the experiment by demonstrating their skill at evaluating their teammates. An inconsistent evaluator would be perceived as inaccurate and, in this context, be less attractive than consistent evaluators. A more serious problem for the Deutsch and Solomon study is Skolnick's (1971) failure to replicate their results even though he used an almost identical procedure. In fact, Skolnick found support for the prediction from self-esteem theory; low self-evaluators rated the positive note sender more favorably and the negative note sender less favorably than did high self-evaluators.

A third experiment tending to support the self-consistency prediction was reported by Dutton and Arrowood (1971). Subjects were led to believe that they had performed either well or poorly in presenting arguments about draft resistance to a panel of peers who, independently of, and crosscutting the performance manipulation, ostensibly judged the subjects' performances as either a "good job" or a "poor job." The results showed that subjects who thought they had done well were more attracted to the peers who made positive evaluations than to those who made negative evaluations. Subjects who thought they had done poorly were more attracted to peers who made negative evaluations.

Although this study appears to support consistency theory, there is reason to question whether or not the good and poor performance conditions represent a self-evaluation manipulation. The task that subjects performed consisted of reading arguments prepared by the experimenter, not the subject. In the poor-performance condition the subject was given an objectively confusing task, then informed by a physiological measure that she felt "extremely anxious," a condition which presumably arose from the "cognition that she was performing poorly." In the good performance condition, the task was objectively clear and the physiological measure indicated no anxiety. Thus, the subject merely read an argument, which was either clear or confusing and which the *experimenter* created, and subsequently her peers either agreed or disagreed with what her body presumably was telling her. Under these circumstances it is hard to see how the subject felt any personal responsibility for either her performance or the consequences of her performance.

Self-Esteem Findings

Both correlational and experimental data have supported self-esteem theory.

In an extensive field investigation of the correlates of self-esteem, Rosenberg (1965) asked his sample of high school juniors and seniors, "How much does it bother you to find that someone has a poor opinion of you?" Since Rosenberg correlated answers to this question with a range of self-esteem scores, this aspect of his study partially fulfills the design criteria for this review. In effect, high and low self-evaluators react to imagined negative evaluations from others. In support of the prediction from self-esteem theory, Rosenberg found that the lower a person's self-esteem, the more he was bothered by the poor opinion of another person.

The earliest experimental study of the relation between self-esteem and interpersonal attraction based on the self-esteem formulation was conducted by Dittes (1959). Self-esteem was measured by various self-reports made by the subjects and by the actual evaluations made by other members of their groups. At the end of a group discussion, the members were permitted to see experimentally manipulated peer ratings of their acceptability to the group. Dittes found that acceptance by the group interacted significantly with self-esteem. For subjects who learned that they were not acceptable to the group, the lower their self-esteem, the less attracted they were to the group. However, for subjects learning that they were acceptable to the group, there was a slight tendency for those with lower self-esteem to be attracted more to the group. A similar interaction effect was obtained by Jones, Knurek, and Regan (in press). In their experiment self-esteem was measured with the Janis-Field scale (Hovland & Janis, 1959). Following a half-hour period for becoming acquainted with another person, subjects learned that their partners either liked them or disliked them. On both a "desire to change partner" measure an a "general emotional reaction" measure low self-esteem subjects indicated more attraction to their partners in the like condition and slightly less attraction to their partners in the dislike condition than did high-self-esteem subjects.

Walster and her colleagues have obtained similar results from experiments in which self-esteem was manipulated rather than measured. They provided subjects with false feedback from personality tests indicating either that they were mature, strong, competent, socially responsible, etc. (high-self-esteem conditions) or immature, inflexible, and possessing other negative characteristics (low-self-esteem conditions). In the first study, Walster (1965) found that female subjects in the low-self-esteem condition were more attracted to a male suitor than were subjects in the high-self-esteem condition. In the second study (Jacobs, Berscheid, & Walster, 1971), the dating skills of a male subject were either praised and he was accepted as a potential date by a same-aged female, or his dating skills were derogated and he was rejected as a potential date. Once again, males in the low-self-esteem condition liked the female evaluator more in the accepting condition but also disliked her more in the rejecting condition than did males in the high-self-esteem condition.

Results supporting the self-esteem theory have been obtained not only for measures and manipulations of global self-esteem but also for more specific aspects of the self. In addition to Skolnick's study (1971) already cited, Jones (1966) manipulated subjects' evaluations of their design judgment ability and subsequently had groups of subjects exchange evaluations of performances relating to this ability. Half of the subjects received mostly positive evaluations from the group, and the remainder received mostly negative evaluations. The evaluations subjects made of their own and other members' performances constituted the main dependent variables. Although the pattern of data indicating subjects' ratings of their peers' performances were in the direction predicted by self-esteem theory, the interaction between manipulated self-evaluation and evaluations received from others was not statistically significant. However, Jones correlated the subjects' evaluations of their own performance with those made of other members' performances and found results supporting the self-esteem theory. Specifically, in the positive evaluation condition, the more positively subjects rated their own performances, the more negatively they rated the responses of others, but in the negative evaluation condition, the correlation between ratings of self and others were positive. Although this latter correlation was not significant, it differed significantly from the corresponding correlation in the positive evaluation condition.

Regarding these correlational findings, it is interesting to note that Harvey (1962) obtained comparable results in a study which generally is viewed as support for the self-consistency theory. His main finding was that the more negative the evaluations a subject received from another person, the less favorable were his attitudes toward the other person. These results can be explained by either self-consistency or self-esteem theory. However, in an internal correlational analysis, Harvey found that the more a subject changed his evaluation of himself in a negative direction, the more he also changed his evaluation of the other person in a negative direction. Since the experimentally controlled evaluations subjects received from the other person were in four out of five conditions negatively discrepant from his self ratings, this positive correlation contradicts the self-consistency theory. This theory would predict that as the self-evaluation changes in the negative direction advocated by the other person, the attraction to this other person will increase.

Secord and Backman's (1961) version of self-consistency theory views the self-evaluation as more flexible and subject to change than most other versions of the theory. Consequently, studies apparently supporting self-esteem theory but measuring only the evaluations a subject makes of others and not changes in his self-evaluation must be interpreted with some caution. In this light the findings of Jones and of Harvey, which had evaluations of both self and others as dependent variables, become particularly important in a comparative evaluation of self-esteem and self-consistency theories.

One further study relevant in this issue of the correlations between evaluations of self and others was conducted by Jones, Hester, Farina, and Davis (1959). Self-evaluation was measured indirectly by asking subjects to indicate how they expected to be rated by two observers on such dimensions as self-centeredness, popularity, tolerance, generosity, and self-confidence. The data reported by Jones and his colleagues concerned subjects' evaluations of the likability of an observer who evaluated them in a highly derogatory fashion on the same dimensions for which subjects had predicted the observers' reactions. The characteristics of the derogator were manipulated. In one condition the derogator was portrayed as unstable and maladjusted and in the other condition as well adjusted and insightful.

Before viewing the data, consider the predictions deriving from the two theories. As noted above, when the evaluations received from the other person are negative, as they were in the present study, the self-consistency theory prediction is that the lower the self-evaluation (in this case the expected ratings of others), the more positive the evaluation of others (i.e., a negative correlation). The self-esteem prediction is that the lower the self-evaluation, the more negative the evaluation of others (i.e., a positive correlation). Furthermore, assuming that the appraisals received from the well-adjusted derogator are more important than those received from the maladjusted derogator, then, on the one hand, from self-consistency theory one would expect a more negative correlation in the adjusted other condition than the maladjusted other condition. The evaluations of the adjusted other presumably are more accurate, and consistency with regard to self-ratings is correspondingly more important. On the other hand, from self-esteem theory one would expect a more positive correlation in the adjusted other condition than the maladjusted other condition. The derogation of the well-adjusted peer is more damaging to self-esteem needs and should have more negative impact on the person with low expectancies about his value for others.

TABLE 2

SUMMARY OF STUDIES

Author(s)	Status of evaluations from others	Theory supported
Backman & Secord (1962)	Estimated by subject	Consistency
Deutsch & Solomon (1959)	Controlled by experimenter	Consistency
Dittes (1959)	Controlled by experimenter	Esteem
Dutton & Arrowood (1971)	Controlled by experimenter	Consistency
Harvey (1962)	Controlled by experimenter	Esteem
Jacobs, Berscheid, & Walster (1971)	Controlled by experimenter	Esteem
Jones, Hester, Farina, & Davis (1959)	Controlled by experimenter	Esteem
Jones (1966)	Controlled by experimenter	Esteem
Jones, Knurek, & Regan (*in press*)	Controlled by experimenter	Esteem
Newcomb (1956)	Estimated by subject	Consistency
Potter (1970)	Controlled by experimenter	Esteem
Rosenberg (1965)	Controlled by experimenter	Esteem
Skolnick (1971)	Controlled by experimenter	Esteem
Walster (1965)	Controlled by experimenter	Esteem
Wiest (1965)	Estimated by subject	Consistency
Wilson (1965)	Controlled by experimenter	Consistency

The correlational data support the self-esteem analysis. First, in the adjusted derogator condition, the relation between ratings of self and other is highly positive (.84) and corresponds in direction to the results obtained by Jones (1966) and Harvey (1962). Second, this correlation is significantly more positive than in the maladjusted derogator condition where, in fact, the correlation is strongly negative (−.62).

One final piece of supporting evidence for the self-esteem position comes from a study by Potter[2] who manipulated subjects' appraisals of their own social sensitivity as well as whether or not they were liked by their partners. Although subjects were more attracted to their partners when their partners liked them as compared to when their partners disliked them, this tendency was significantly greater for low-self-evaluators than for high-self-evaluators.

---

[2] D. A. Potter. Accuracy and interpersonal attraction. Unpublished doctoral dissertation, Cornell University, 1970.

Summary of Evidence

The implications of these studies, which may be viewed as "critical tests" between the two theories, are summarized in Table 2. The table also summarizes the status of the evaluations subjects experienced from others—that is, whether these evaluations were estimated by the subject or experimentally controlled. Two aspects of the table are worth noting. First, the evidence, in general, tends to favor self-esteem theory over self-consistency theory. Of the 16 investigations reviewed, 10 support self-esteem theory, and the author has suggested that there are serious problems of interpretation or replication with the experimental studies often cited as support for self-consistency theory. Furthermore, the self-esteem evidence becomes more impressive when one considers the variety of different procedures and measures used in these studies. However, at this point, to draw any firm conclusions about the relative value of self-esteem theory would be premature. Some data do seem to support self-consistency theory, and the question is whether these data can be accounted for within an expanded self-esteem theory framework.

A second feature of Table 2 is that all three studies in which peer evaluations were estimated by the subject support self-consistency theory, whereas 10 of the 13 studies in which these evaluations were in some way controlled by the experimenter support self-esteem theory. This pattern of results suggests that a subject may be confronted with quite different problems depending upon whether he must estimate another person's feelings or is actually informed of those feelings as they relate to some aspect of himself. This difference is discussed later on in the paper.

TOWARD A RESOLUTION

People often comment on the "counter-intuitive" quality of many predictions from cognitive consistency theories, and some of the implications of self-consistency theory are subject to this general bafflement. Many authors working primarily within a cognitive consistency framework have acknowledged the discrepancies between cognitive consistency motivations and self-esteem, self-enhancement, or even achievement motivations, and have expressed such reservations. Rosenberg (1968), for example, observed that one of the most important failings of the consistency approach

> seems to be the experimental affirmation of this hard, intractable fact: the need for (or trend toward) the maintenance of internally consistent affective cognitive structures is often subordinated to man's penchant for trying to think well of himself and optimistically of his prospects [p. 384]."

In response to this dilemma, some authors have viewed esteem motivations as "in addition to" forces affecting behavior in experiments designed to demonstrate self-consistent responses (e.g., Aronson and Carlsmith, 1962; Pepitone, 1964) whereas others have tried to expand theories of consistency to account for self-esteem tendencies (e.g., Secord & Backman, 1965).

The author's own orientation to this problem was to ask whether apparently self-consistent actions and attitudes might not be explicable solely in terms of self-esteem theory. In particular, two variables which may be viewed as extensions of this latter theory—the anticipation of future consequences from accepting or rejecting the evaluations of others, and the degree of personalistic focus attributable to these evaluations—proved to be useful in accounting for a variety of results in this area of self- and interpersonal evaluations.

Consequences of Accepting or Rejecting Evaluations

As one alternative interpretation of self-consistent responses, consider the possibility that they reflect a tendency for people to react to evaluative information in terms not only of its present but also its future implications for esteem enhancement. In general, people prefer to accept approval from others and to reject disapproval. However, there are numerous occasions on which a person's accepting inconsistent and inaccurate evaluations for himself has important and potentially hazardous consequences for his future decisions, responsibilities, and performances, and these consequences can be as frightening for the low-self-evaluator who is overrated as they are humiliating for the high-self-evaluator who is underrated. In addition to the failures and limited successes of his actual performances, the individual also may suffer social embarrassment from misrepresenting himself or from the exposed discrepancy between others' expectancies and his actual effectiveness.

These anticipated successes or failures and expressions of approval or disapproval which are contingent on the individual's accepting evaluations of himself can be satisfying or frustrating to his need for self-esteem enhancement. Hence, this position assumes that some apparently self-consistent responses can be viewed in terms of people forfeiting immediate gratifications of esteem needs in anticipation of more self-enhancing or less self-derogating experiences in the future.

Two classes of conditions have been identified which are assumed to increase the salience of the future consequences of interpersonal evaluations: *predecisional conditions* in which a person anticipates making choices for himself which involve his self-evaluation, and *self-exposure conditions* in which a person anticipates the public revelation of his actual successes or failures. The general hypothesis guiding the research described below is that the anticipation of making a personal decision or of self-exposure will increase the tendency of the individual to make apparently self-consistent responses in his relations with others.

In a study of the effects on interpersonal evaluations of anticipating personal decisions, Jones and Ratner (1967) induced all subjects to adopt negative evaluations of their "personality inference" ability by giving them false low scores on each of a series of four personality tests. In a second part of the experiment three subjects at a time participated in a group session in which they evaluated each other's responses to items like those appearing on the personality tests. The evaluations from peers were experimentally controlled so that each subject received mostly positive evaluations from one peer and mostly negative evaluations from the other. Before actually beginning this second phase of the experiment, subjects were told that there would be a third part of the study in which they would be required to use this "personality inference" ability in a case analysis which would be evaluated by a clinical psychologist. Half the subjects immediately chose how difficult a clinical case they wished to work on (postdecision condition), and, as expected, their choices were from among the easier cases. The remainder of the subjects anticipated making their choices after the group session (predicision condition).

The argument made by Jones and Ratner was that prior to making a decision for himself, a subject would prefer accurate, self-consistent information from others, but that after making the decision as to clinical task he would, in effect, be protected from the hazardous consequences of accepting erroneous evaluations and would prefer information providing more esteem gratifications, that is, positive evaluations. The results showed that in the postdecision condition subjects evaluated the responses of the positive evaluator more favorably than the negative evaluator but that in the predecision condition subjects more favorably evaluated the responses of the negative evaluator.

Three experimental investigations have been conducted which relate to the anticipation of self-exposure. A study by Jones and Pines (1968) used a procedure very similar to the study just described. Once again, subjects were given negative feedback on the "personality inference" tests and, subsequently, during a group session received mostly positive evaluations from one peer and mostly negative evaluations from the other. One further manipulation involved self-exposure. Half the subjects were led to believe that following the group session they would once again give their answers aloud but the correctness of their answer would be indicated publicly by the experimenter. Thus, in this condition subjects expected to have their level of skill (which they thought was low) revealed to their peers. The other half of the subjects were not led to expect such self-exposure.

As predicted and in support of the theoretical analysis presented above, subjects expecting self-exposure increased the favorability of their appraisals of the negative evaluator relative to the positive evaluator as the time of the experimenter's public evaluation approached, and those not expecting self-exposure became increasingly more favorable to the positive evaluator. Furthermore, subjects in the self-exposure condition reported that the negative test evaluation from experts was a more meaningful indication of their abilities, whereas those in the no self-exposure condition found the relatively more positive evaluations of the peers (in combination) were the better indicators of their ability.

An unpublished dissertation by Haraguchi[3] was a conceptual replication of the Jones-Pines study but employed a quite different procedure suggested by Jones' (1964) analysis of self-presentation. Subjects were led to believe from a highly qualified source that they would do relatively poorly on an achievement test. However, the subsequent estimate of an inexperienced clinical interviewer was that they would do very well on the test. The prediction was that subjects would present themselves in a more modest fashion to the interviewer under self-exposure conditions in which they expected to discuss their test scores with the interviewer as compared to the no-self-exposure conditions.

Although the main analysis of the self-presentation data revealed no significant effect of the self-exposure variable, Haraguchi found in an internal analysis that the results were very much affected by the perceived attractiveness of the interviewer. She reasoned that subjects would be more likely to avoid embarrassment and disapproval to the extent that they favorably evaluated their interviewers. The results of this internal analysis showed that for the half of the subjects who were most attracted to their interviewers the results were in accordance with her initial prediction; that is, those subjects in the self-exposure condition evaluated themselves less favorably (and more consistently with respect to their self-evaluations) than those in the no-self-exposure conditions. These results were significantly reversed for the remainder of the subjects who were least attracted to their interviewers.

Finally, Eagly and Acksen (1971) conducted a study which may be viewed within the present framework. In their experiment subjects received notes from trained graduate students indicating that they were either more creative or less creative than they had evaluated themselves. Subsequently, changes in self-appraisals were measured either under two conditions in which subjects expected to take a creativity test in which their degree of success would be obvious to themselves and the graduate student administering the test, or under a condition in which no tests were expected. The authors obtained an interaction between expectancy of testing and favorability or unfavorability of the student's note. Only when subjects received a favorable note and expected no test did they change their self-evaluation in the direction of being more creative. Furthermore, in the favorable note conditions subjects changed their self-evaluations more in the direction of the message when they anticipated no test as compared to when they did expect the test; the corresponding comparison for the unfavorable note conditions revealed no significant differences.

In general, the results reported in this section support the notion that apparently self-consistent behavior may result from the future implication of evaluations for esteem-enhancement. Whether people enhance self-esteem by accepting positive evaluations of themselves or by being more realistic about themselves has been shown to be affected by their expectancies regarding events which could reveal their true worth.[4]

Implications of "Personalism" for Self-Esteem Enhancement

There is a second analysis which suggests that self-consistent interpersonal evaluations may be interpreted with the assumptions of self-esteem theory. Given that an individual is liked, approved of, or agreed with by another person, he probably finds this information more rewarding to the extent that its cause is attributable to characteristics of himself rather than to some cause which is unrelated to himself. For example, thinking "He likes me because of the kind of person I am." is undoubtedly more rewarding than thinking "He likes me because he's the kind of person who likes everybody." By the same token various sorts of negative evaluations may well be increasingly more damaging to self-esteem needs when the individual sees himself as the cause of the evaluations as compared to when the cause is externally attributed.

---

[3] R. S. Haraguchi. Interpersonal conditions affecting self presentation. Unpublished doctoral dissertation, State University of New York at Buffalo, 1967.

[4] Jones (1968a) and Mettee (1971) have made a similar argument in an attempt to explain apparently self-consistent responses in a nonsocial testing situation designed by Aronson and Carlsmith (1962).

This analysis draws heavily upon Jones and Davis' (1965) concept of personalism. In line with their thinking, I assume that to the extent that actions (such as interpersonal evaluations) are attributable to, or "uniquely focused" on the self, those actions have more impact on the self-esteem needs of the individual. Thus, to the degree that positive evaluations are perceived as focusing on the self, esteem needs are satisfied; and to the degree that negative evaluations are focused in this manner, esteem needs are frustrated.

An important variable affecting personalistic attributions may well be the similarity or consistency existing between the individual's evaluation of himself and the evaluations he receives from the actor. When the individual and another party have similar opinions about an entity, there is at least the possibility that the other person is focusing on, and being directly affected by the same events as the subject in question (cf. Kelley, 1967).

If self-consistency produces the attribution of personal cause, then this analysis, once again, leads to the conclusion that self-consistency per se is not a general motivational state. Rather, personalism via self-consistency interacts with the esteem enhancing or derogating value of the actions directed toward the individual. Thus, the hypothesis to be considered is that when an individual is accepted or liked by others, the more self-consistent their evaluations of his characteristics, the more he is attracted to them, but when an individual is rejected or disliked by others, the more self-consistent their evaluations, the less he is attracted to them.

Two studies were designed to test this general notion. In the first study, Jones et al. (in press) assumed (and self-report data confirmed this assumption) that subjects would have generally favorable ratings of their own characteristics and manipulated self-consistency by providing the subjects with positive (and consistent) trait ratings by their partners or with negative (and inconsistent) ratings. Subsequently, whether a partner liked or disliked a subject was varied, and measures of the subject's attraction to his partner revealed a significant interaction between self-consistency and liking which partially supported the above hypothesis. When subjects were liked by their partners, their attraction was greater when they received positive trait ratings rather than negative ratings from their partners; when subjects were disliked they were slightly less attracted when they received positive ratings. Although alternative interpretations of these findings were implausible in the light of other data obtained in this study, a more substantial test of the personalism hypothesis requires a direct manipulation of self-consistency.

An unpublished dissertation by Potter (see Footnote 2) permits such a test. Subjects were first given two written tests of their social sensitivity and received test feedback indicating that they had either a high level of skill or a low level of skill. Then they participated in a structured "social sensitivity interview" in the presence of a partner who subsequently evaluated them as being either well above average or well below average in social sensitivity. Since there was evidence that the test feedback affected self-evaluations, the test and interpersonal evaluations of social sensitivity, which constituted a 2×2 design may be considered in conjunction as a manipulation of self-consistency. Evaluations which were similarly high or low are consistent, but evaluations which were dissimilar are inconsistent. Finally, on an additional evaluation form subjects were informed that their partners either liked or disliked them.

TABLE 3

Attraction to Partners as a Function of Self-Consistency and Liking

Self-consistency	Like[a]	Dislike
Consistent	2.34	4.65
Inconsistent	3.06	3.44

[a] The lower the number, the greater the subject's liking for his partner.

The major dependent variable of this study was a question designed to assess how much a subject liked his partner. If self-consistency is a general motivational tendency, then one would expect that regardless of whether they were liked or disliked by their partners, subjects would be more attracted to them in the consistent conditions than the inconsistent conditions. If, on the other hand, the analysis of personalism is correct, subjects in the like condition would be attracted to partners in the consistent condition more than partners in the inconsistent condition, but subjects in the dislike condition would like consistent partners less than inconsistent partners. As shown in Table 3, the pattern of results is precisely as predicted by the personalism analysis, and the interaction between self-consistency and liking is highly significant ($p < .001$).

The results of these studies provide support for the assumption that personalism via self-consistency increases the impact of others' feelings on the individual's self-esteem needs. People do prefer self-consistent evaluations from others when they feel that they are accepted or liked by others. However, the fact that inconsistent evaluations are preferred when the individual finds himself disliked strongly indicates that esteem enhancement and not maintenance of stable cognitions is the motivational force operating in these social relations.

This analysis of self-consistency has implications for two of the correlational studies reported earlier in the paper. Backman and Secord (1962) and Newcomb (1956) showed that an individual's liking for others is a direct function of the similarity between his evaluations of himself and those he attributes to others. Although these findings seem to support self-consistency theory, a self-esteem interpretation now becomes possible. To make this interpretation meaningful it is necessary to consider how much the individual felt he was liked by each of his peers. In these survey studies of ongoing relations, it is probably safe to assume that the individual's liking for others is highly correlated with how much others like him or at least with his perception of their liking for him (see, e.g., Tagiuri, Bruner, & Blake, 1958, for evidence supporting this assumption). From this assumption follows the possibility that the Backman and Secord, and Newcomb findings indicate that the more the individual likes and thinks he is liked by another person, the more he attributes similarity with regard to opinions about himself. This tendency falls into line with the pattern of results found by Potter. It suggests that self-consistent opinions are attributed to friends to enhance the value of their affection whereas dissimilar opinions tend to be attributed to enemies to defend against the sting of their rejection.

"Warm" and "Cool" Studies of Social Evaluation

The analyses of future consequences and personalism suggest the conclusion that a variety of self-consistent results may be explained within the framework of self-esteem theory and that cognitive consistency motivations play at best a minor role in interpersonal evaluations of the sort discussed here. However, there may be limitations to the generalizability of this conclusion to some kinds of investigations of social perception and evaluation. There are a number of studies (e.g., Aronson & Cope, 1968; Jordan, 1953) showing that people tend to hold consistent or balanced cognitions with respect to their evaluations of others. Indeed the correlational studies reviewed earlier (Backman & Secord, 1962; Newcomb, 1956; and Wiest, 1965), although open to alternative interpretations, show similar consistency effects. A major distinction which may be useful in those studies which obtain results supporting cognitive consistency theory from those supporting self-esteem theory involves whether or not the experimental subjects are themselves targets of evaluative action from the stimulus person being evaluated. Experiments in which the subject is liked or disliked, agreed or disagreed with, benefited or harmed, or in other ways directly affected by the stimulus person may be considered "warm" studies. Since these subjects are to some extent the focus of actions having evaluative implications, such studies should be interpretable within the self-esteem theory framework. However, experiments in which the subject is in no way personally affected by the stimulus person may be viewed as "cool" studies. In the literature on person perception there are many such investigations in which the subject observes but does not interact with some staged slice of life involving other people, then is asked to form impressions of these others based on his observations. Since there is no direct personal involvement and the orientation of the subject is essentially clinical or diagnostic, these studies may well be more explicable in terms of cognitive consistency theory.

This distinction is helpful in accounting for the pattern of results reported previously in Table 2. Those studies in which the subject estimates the reactions of others are relatively cool because the subject is neither interacting with, nor receiving evaluative information from others. Hence, his esteem needs are of less concern to him than the cognitive task of making sense of the social world he watches, and, indeed, his attitudes in these situations do tend to be self-consistent. Those studies in which the subject receives evaluations from others are warm studies. The subject learns what others actually think of him, he worries about his own esteem and worth, and, as the data indicate in most cases, he is differentially satisfied or frustrated depending upon others' evaluations and his self-evaluation.

The variable differentiating warm and cool studies—whether or not the subject is the focus of evaluative actions from the target person—is easily manipulable, and the idea that warm situations will lead to esteem theory results whereas cool situations will lead to consistency theory results should be experimentally tested. In addition, systematically applying this distinction to other studies in person perception should be undertaken before drawing any firm conclusions. If the difference proposed turns out to be meaningful, there is an interesting implication for the person perception literature in general; namely, the popularity of consistency theories in this area may be in part at least a function of the many studies conducted in which the subject is required to comment upon or evaluate a social game that he observes from the sidelines but in which he rarely participates. Furthermore, if living with others is a matter of playing the game rather than watching it, then the generality of consistency models to real life situations may be quite limited.

SUMMARY AND CONCLUSIONS

The purpose of this paper was to compare two rather broad theoretical traditions in social psychology as they related to the determinants of interpersonal evaluations. On the one hand, proponents of self-consistency theory argue that the individual adjusts his cognitions and orients his relations with others so as to maintain similarity between his evaluations of himself and those he receives from others. On the other hand, the assumptions of self-esteem theory are that the individual has a need for self-esteem which is satisfied primarily by the approval he receives from others and is frustrated by their disapproval, and that to the extent that his esteem need is satisfied by others' evaluations, the individual will respond favorably to them.

In spite of the fact that the cognitive consistency approach has been widely accepted as a theoretical framework for these problems as well as many other problems in social psychology, I suggest that a stronger case can be made for the value of the self-esteem theory outlined in this paper. The evidence from experiments which permit a test between these positions provides more substantial support for the self-esteem predictions. Time and again, the unhappy self-derogator seems to glow when praised and glare when censured even more than his self-confident counterpart. Furthermore, two rather straightforward extensions of self-esteem theory are capable of explaining apparently self-consistent interpersonal evaluations. First, the anticipated exposure of his qualities tends to make the individual more "honest" about himself (and, therefore, more self-consistent) so as to avoid the eventual disapproval of others. Second, people respond more favorably to positive evaluations and less favorably to negative evaluations to the extent that those evaluations are perceived as uniquely focused on themselves. Not only can these propositions account for some of the data often cited as support for a self-consistency motivation, but they also point in the direction of sorting out self-consistent responses from other strategies designed to gain esteem enhancement.

Whether or not these conclusions have implications for other social psychological problems involving cognitive consistency models remains to be seen. Bramel (1968) has recently suggested that certain cognitive dissonance phenomena may reflect underlying needs for self-esteem enhancement through social approval. He argues that "dissonance is a feeling of personal unworthiness (a type of anxiety) traceable to rejection of oneself by other people either in the present or in the past" and that the "consequences of anxiety about one's worth are likely to be such things as self-justification and the search for information that will reflect favorably upon the self [p. 365]." Deutsch, Krauss, and Rosenau (1962) have made a similar argument. Exploring the similarities and differences among these various lines of analysis could lead to a more clearly articulated self-esteem theory as well as sharpen the distinction between it and cognitive consistency theories.

EPILOGUE

I am good
You love me
therefore you are good.
I am bad
You love me
therefore you are truly beautiful!

REFERENCES

ARONSON, E. Some antecedents of interpersonal attraction. In W. J. Arnold & D. Levine (Eds.), *Nebraska symposium on motivation.* Lincoln, Nebraska: University of Nebraska Press, 1969.

ARONSON, E., & CARLSMITH, J. M. Performance expectancy as a determinant of actual performance. *Journal of Abnormal and Social Psychology,* 1962, **65**, 178–182.

ARONSON, E., & COPE, V. My enemy's enemy is my friend. *Journal of Personality and Social Psychology,* 1968, **8**, 8–12.

ARONSON, E., & METTEE, D. R. Dishonest behavior as a function of differential levels of induced self-esteem. *Journal of Personality and Social Psychology,* 1968, **9**, 121–127.

ARONSON, E., & WORCHEL, P. Similarity vs. liking as determinants of interpersonal attractiveness. *Psychonomic Science,* 1966, **5**, 157–158.

BACKMAN, C. W., & SECORD, P. F. The effect of perceived liking on interpersonal attraction. *Human Relations,* 1959, **12**, 379–384.

BACKMAN, C. W., & SECORD, P. F. Liking, selective interaction, and misperception in congruent interpersonal relations. *Sociometry,* 1962, **25**, 321–335.

BARON, R. M. Social reinforcement effects as a function of social reinforcement history. *Psychological Review,* 1966, **73**, 527–539.

BAVELAS, A. HASTORF, A. H., GROSS, A. E., & KITE, W. R. Experiments on the alteration of group structure. *Journal of Experimental Social Psychology,* 1965, **1**, 55–70.

BRAMEL, D. Dissonance, expectation, and the self. In Abelson et al. (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

CARTWRIGHT, D., & ZANDER, A. *Group dynamics.* (2nd ed.) Evanston, Ill.: Row, Peterson, 1960.

DEUTSCH, M., KRAUSS, R. M., & ROSENAU, N. Dissonance or defensiveness? *Journal of Personality,* 1962, **30**, 16–28.

DEUTSCH, M., & SOLOMON, L. Reactions to evaluations by others as influenced by self-evaluation. *Sociometry,* 1959, **22**, 92–113.

DITTES, J. E. Attractiveness of group as a function of self esteem and acceptance by group. *Journal of Abnormal and Social Psychology,* 1959, **59**, 77–82.

DOLLARD, J., DOOB, L., MILLER, N. E., MOWRER, O. H., & SEARS, R. R. *Frustration and aggression.* New Haven: Yale University Press, 1939.

DUTTON, D. G., & ARROWOOD, A. J. Situational factors in evaluation congruency and interpersonal attraction. *Journal of Personality and Social Psychology,* 1971, **18**, 222–229.

EAGLY, A. H., & ACKSEN, B. A. The effect of expecting to be evaluated on change toward favorable and unfavorable information about oneself. *Sociometry,* 1971, **34**, 411–422.

FESTINGER, L. A theory of social comparison processes. *Human Relations,* 1954, **7**, 117–150.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson, 1957.

HARVEY, O. J. Personality factors in resolution of conceptual incongruities. *Sociometry,* 1962, **25**, 336–352.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

HOMANS, G. C. *Social behavior: Its elementary forms.* New York: Harcourt, Brace & World, 1961.

Hovland, C. I., & Janis, I. L. (Ed.) *Personality and persuasibility*. New Haven: Yale University Press, 1959.

Jacobs, L., Berscheid, E., & Walster, E. Self-esteem and attraction. *Journal of Personality and Social Psychology*, 1971, **17**, 84–91.

Jones, E. E. *Ingratiation: A social psychological analysis*. New York: Appleton-Century-Crofts, 1964.

Jones, E. E., & Davis, K. E. From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. New York: Academic Press, 1965, Vol. 2.

Jones, E. E., & Gerard, H. B. *Foundations of social psychology*. New York: Wiley, 1967.

Jones, E. E., Hester, S. L., Farina, A., & Davis, K. E. Reactions to unfavorable personal evaluations as a function of the evaluator's perceived adjustment. *Journal of Abnormal and Social Psychology*, 1959, **59**, 363–370.

Jones, S. C. Some determinants of interpersonal evaluating behavior. *Journal of Personality and Social Psychology*, 1966, **3**, 397–403.

Jones, S. C. Expectation, performance and anticipation of self-revealing events. *Journal of Social Psychology*, 1968, **74**, 189–197. (a)

Jones, S. C. Some effects of interpersonal evaluations on group process and social perception. *Sociometry*, 1968, **31**, 150–161. (b)

Jones, S. C., Knurek, D. A., & Regan, D. J. Variables, affecting reactions to social acceptance and rejection. *Journal of Social Psychology*, in press.

Jones, S. C., & Panitch, D. The self-fulfilling prophecy and interpersonal attraction. *Journal of Experimental Social Psychology*, 1971, **7**, 356–366.

Jones, S. C., & Pines, H. A. Self-revealing events and interpersonal evaluations. *Journal of Personality and Social Psychology*, 1968, **8**, 277–281.

Jones, S. C., & Ratner, C. Commitment to self-appraisal and interpersonal evaluations. *Journal of Personality and Social Psychology*, 1967, **6**, 442–447.

Jordan, H. Behavioral forces that are a function of attitudes and of cognitive organization. *Human Relations*, 1953, **6**, 273–287.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation, 1967* Lincoln, Nebraska: University of Nebraska Press, 1967.

Laing, R. D. *Knots*. New York: Pantheon, 1970.

Lecky, P. *Self-consistency: A theory of personality*. New York: Island Press, 1945.

Mettee, D. R. Rejection of unexpected success as a function of the negative consequences of accepting success. *Journal of Personality and Social Psychology*, 1971, **17**, 332–342.

Morse, S., & Gergen, K. J. Social comparison, self-consistency, and the concept of self. *Journal of Personality and Social Psychology*, 1970, **16**, 148–156.

Newcomb, T. M. The prediction of interpersonal attraction. *American Psychologist*, 1956, **11**, 575–586.

Newcomb, T. M. *The acquaintance process*. New York: Holt, Rinehart & Winston, 1961.

Pepitone, A. *Attraction and hostility*. New York: Atherton Press, 1964.

Pepitone, A. Some conceptual and empirical problems of consistency models. In Feldman (Ed.), *Cognitive consistency*. New York: Academic Press, 1966.

Rogers, C. R. *Client-centered therapy*. Boston: Houghton, 1951.

Rogers, C. R. A theory of therapy, personality, and interpersonal relationships as developed in the client-centered framework. In S. Koch (Ed.), *Psychology: A study of science*. Vol. 3. New York: McGraw-Hill, 1959.

Rosenberg, M. *Society and the adolescent self-image*. Princeton, N. J.: Princeton University Press, 1965.

Rosenberg, M. J. Discussion: The concepts of self. In Abelson et al. (Eds.), *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally, 1968.

Secord, P. F., & Backman, C. W. Personality theory and the problem of stability and change in individual behavior: An interpersonal approach. *Psychological Review*, 1961, **68**, 21–32.

Secord, P. F., & Backman, C. W. *Social psychology*. New York: McGraw-Hill, 1964.

Secord, P. F., & Backman, C. W. An interpersonal approach to personality. In B. A. Maher (Ed.), *Progress in experimental personality research*. Vol. 2. New York: Academic Press, 1965.

Skolnick, P. Reactions to personal evaluations: A failure to replicate. *Journal of Personality and Social Psychology*, 1971, **18**, 62–67.

Tagiuri, R., Bruner, J. S., & Blake, R. R. On the relation between feelings and perception of feelings among members of small groups. In E. E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, Rinehart & Winston, 1958.

Thibaut, J. W., & Kelley, H. H. *The social psychology of groups*. New York: Wiley, 1959.

Videbeck, R. Self-conception and the reactions of others. *Sociometry*, 1960, **23**, 351–359.

Walster, E. The effect of self-esteem on romantic liking. *Journal of Experimental and Social Psychology*, 1965, **1**, 184–197.

Wiest, W. M. A quantitative extension of Heider's theory of cognitive balance applied to interpersonal perception and self-esteem. *Psychological Monograph*, 1965, **79**(14, Whole No. 607).

Wilson, D. T. Ability evaluation, postdecision dissonance, and coworker attractiveness. *Journal of Personality and Social Psychology*, 1965, **1**, 486–489.

(Received November 26, 1971)

Psychological Bulletin
1973, Vol. 79, No. 3, 200–210

COGNITIVE FACTORS IN ELECTRODERMAL CONDITIONING[1]

WILLIAM W. GRINGS[2]

University of Southern California

An analysis is made of concepts central to explanation of cognitive determination of electrodermal changes in conditioning situations. Past issues have emphasized verbally induced task sets, awareness of stimulus contingencies, and the possibility of simultaneous learning at different "levels." Current issues emphasize verbally determined changes in the perceptual significance of stimuli and changes in perception of relations among stimulus events. In conditioning situations, electrodermal activity is seen as a sensitive indicator of the momentary modulations of arousal which accompany processes described by the terms of orientation, anticipation, expectation, preparation, and signal-induced adaptation or defense.

After numerous early studies demonstrated the feasibility of conditioning the electrodermal response, that response was adopted enthusiastically for studies of conditioning because it was assumed to be involuntary and free from the influence of subject verbalizations. By the time of publication of Hilgard and Marquis' (1940) summary of the field, a large number of traditional phenomena had been demonstrated with the electrodermal response. To mention a few, they included inhibition of delay and disinhibition, comparison of trace and delay paradigms, stimulus generalization, spontaneous recovery, partial reinforcement effects, compound stimulus patterning, and semantic generalization.

During the same period, there was an increasing concern for the fact that at least some of the observed results reflected phenomena more like verbal expectancy learning (Cook & Harris, 1937) or preparatory (cognitive) sets for receipt of stimulation (Mowrer, 1938) than phenomena of "true conditioning." As efforts were made to replicate and extend upon the earlier findings, researchers on electrodermal conditioning found themselves in the middle of a number of theoretical–methodological issues involving cognition for the electrodermal response is so sensitive to shifts of attention and the momentary changes in arousal which accompany changes in cognition that it is impossible to eliminate cognitive factors completely from conditioning observational situations. At the same time, this sensitivity gives the electrodermal response an advantage when it comes to direct study of cognitive variables.

The present material is an examination of the role the electrodermal response has played in the study of cognitive factors in conditioning. It extends upon an earlier discussion (Grings, 1965). An arbitrary selection and classification of researches is made as a basis for examination of a few central issues. Where a choice of research paradigms is required, the paired-stimulation arrangement of classical conditioning is given preference over paradigms of avoidance or escape learning, instrumental reward learning, and operant conditioning. The restrictions are not made to suggest a discontinuity among kinds of learning situations, but rather to gain simplicity of explanation for a limited range of phenomena. It is also true that the classical conditioning paradigm has been the most widely researched stimulation arrangement with the electrodermal response.

BACKGROUND FOR CONSIDERATION OF COGNITIVE DETERMINERS OF CONDITIONING

The first clear indications of cognitive learning of autonomic responses were reported over 30 years ago. Cook and Harris (1937) manipulated the electrodermal response to a "neutral" stimulus by instructing subjects that the stimulus would be followed by a shock. They labeled the phenomenon "verbal conditioning" involving "an association between two verbalizations: for example, the verbalization 'light' and the verbalization 'receive a shock' [p. 209]." They asserted further that "conditioning of the galvanic skin response in the human adult differs from the customary conditioning procedure [p. 209]."

---

[1] Adapted from a research lecture presented to the Institute of Psychiatry, University of London, February 11, 1971, while the author was United States National Institute of Mental Health Special Research Fellow (MH 46955). Cited research by the author was supported by National Institute of Mental Health Grant MH 3916.

[2] Reprint requests should be sent to William W. Grings, Department of Psychology, University of Southern California, University Park, Los Angeles, California 90007.

These early observations were put aside by many workers, and electrodermal conditioning continued to be carried out assuming that subjects could be kept from making verbal associations by giving them instructions intended to mislead them about the purpose of an experiment. The results of such verbal factors were considered to be artifacts and not really conditioning (e.g., Mowrer, 1938). One notable exception to this arose from an interest in practical questions of extinguishing undesirable conditioned autonomic responses where conventional extinction procedures were compared to what might be termed "extinction plus knowledge" (e.g., Haggard, 1943).

EARLY STUDIES OF COGNITIVE FACTORS IN EXTINCTION

In the previously cited study, Haggard (1943) investigated a variety of cognitive variables affecting skin resistance phenomena in a situation associating shock with a verbal stimulus. For some subjects the shock was administered by the experimenter; for others it was self-administered. Groups were later divided on whether they were aware of the word–shock relation. Of three types of therapy conditions compared after conditioning, one involved permitting the subject to ask questions, where "an invariable question was whether they would be shocked again" to which the answer "no further shocks" would be given. The results were quite complicated because of the wide variety of variables involved, but several conclusions were evident. One was that cognitive variables have an important effect on general reactivity (motivation, arousal) variables and that the influence of cognition upon extinction, while present, depended upon other variables (like task requirements).

This difference between associative and motivational effects of instructions during extinction of the electrodermal response is characteristic of later studies. Silverman (1960) observed short (.5-second) and long (6-second) interstimulus-interval conditioning groups and a pseudoconditioning control group when half of each group was extinguished with "no more shock" instructions. Knowledge from instructions affected both the .5-second and the pseudoconditioning group performance(produced a decrement) whereas the 6-second group was not affected. The result was interpreted as reflecting anxiety produced by the longer interval.

Wickens, Allen, and Hill (1963) compared extinction with and without such "no more shock" instructions with two levels of unconditioned stimulus (UCS) intensity. The two control groups (instructed and noninstructed) varied in response magnitude showing that one effect of instructions was to lower drive. Instructions speeded extinctions but did not produce it immediately. There was no clear effect of shock intensity. In a related study, Grings and Lockhart (1963) found a performance decrement attributable to instructions but no difference due to UCS intensity. The absence of an unpaired group made it impossible to separate motivational and associative effects of the knowledge about stimulus relations.

Because many investigators use extinction measures as indices of conditioning acquisition, the role of instructions upon extinction can be evaluated in a number of studies of related issues, like those of awareness and levels of learning, to be noted later (e.g., Bridger & Mandel, 1965; Fuhrer & Baer, 1969; Wilson, 1968).

TASK SET

Traditionally one of the most frequent interpretations of cognitive influences in conditioning has been in terms of facilitating or inhibiting effects upon performance through the subject's set toward his task. For eye-blink-conditioning situations these results were extensively discussed (e.g., Hilgard & Humphreys, 1938; Norris & Grant, 1948). The general results have been clear evidence of inhibition and facilitation by instructional variables.

Two recent articles reporting on the electrodermal response illustrate this type of concern. Hill (1967) discussed conditioning with her subjects, then told half of them that "the intelligent thing is to become conditioned," while telling the other half that "the intelligent thing is to not become conditioned." The data revealed differential responding in accordance with instruction, the favorable group responding at a consistently higher level than the antiinstruction group. The author viewed the instruction effect as predominately motivational. Dawson and Reardon (1969) repeated Hill's instructions and added a control group and a neutral instruction group. Both the inhibitory and facilitory group surpassed the control group, but neither differed significantly from a neutral-instruction-conditioning group, although the facilitation group performed significantly better than the inhibitory group. They also found differential unconditioned response (UCR) habituation in the two conditions (inhibiton and facilitation). A later study (Harvey & Wickens, 1971) suggested the possibility that the superior performance of the "pro" group may be due to a nonassociative (drive) variable.

AWARENESS

A frequently researched question during the past two decades concerns the concept of awareness and the role it plays, possibly as a necessary condition to autonomic learning. Put in its strongest form this argument asserts that all autonomic conditioning is verbal expectancy learning and that awareness of conditioned stimulus (CS)–UCS relations is necessary for conditioning to occur. Experiments to that end for both electrodermal response and heart rate conditioning are typified by those of Chatterjee and Ericksen (1960, 1962). Numerous difficulties stand in the way of proving that awareness of CS–UCS relations is necessary for autonomic conditioning to occur. It becomes important to demonstrate an absence of conditioning when there is an absence of awareness. Given the absence of conditioning under unaware circumstances, it is always possible that the other necessary conditions for learning are also not present (like perception of, or attention to, stimuli; and an adequate number of trials for learning to occur). The experiments tend more to prove that cognitive awareness factors are sufficient and potent than that they are necessary.

Some of the difficulties in this type of research can be seen from a brief look at a sample of studies. In early investigations, the variable of awareness was defined by separating subjects into aware and unaware groups on the basis of postexperimental interviews (e.g., Diven, 1937; Lacey, Smith, & Green, 1955). There are many weaknesses in this procedure. Potential sources of bias exist during the interview and the locus of the information collection makes it possible that learning could occur first, then produce the awareness. Another approach collects intertrial reports and attempts to verify the proposition that learning occurs on the trial on which awareness occurs. Such a procedure has been used successfully in studies of verbal conditioning (i.e., nonautonomic) by DeNike (1964), but efforts to use the method with the electrodermal response have not been fully successful (Fuhrer & Baer, 1965; Shean, 1968), requiring the investigators to base their principal arguments upon postexperimental accuracy in reporting CS–UCS relations.

Manipulative definitions of awareness have been of two main types. One, based on differences in instructions about stimulus relations, has already been noted above to enhance performance but to be difficult to evaluate in the opposite direction, that is, in terms of preventing awareness. There is nothing in misleading or disguising instructions to guarantee that a subject will not catch on to the stimulus relations. As a result, a second form of manipulation, the use of a masking task, has recently seen emphasis. In an early evaluation of the effect of a masking task in electrodermal response conditioning, Dawson and Grings (1968) administered conditioning stimuli while subjects were taking a paper-and-pencil test. Their subjects showed no differential response to reinforced and nonreinforced stimuli until they were instructed of stimulus relations. However, the number of trials were few and the possibility existed that subpects were distracted by the task to a degree that might keep them from perceiving the stimuli.

To overcome the weakness just mentioned, Dawson (1970) devised a masking task in which the subject made discriminative judgments among members of a series of tone stimuli, and reinforcement conditions were introduced by pairing shocks with one of the tones. Further manipulation by instructions showed conditioning only when awareness was introduced. In a related study, Fuhrer and Baer (1969) employed differential electrodermal response conditioning in a group who had conditioning embedded in a masking task (probability matching) and a comparison group who underwent conventional conditioning. Content analysis of postexperimental interviews showed that 62% of the subjects in the conventional conditioning group accurately verbalized CS-UCS relations while only 30% of the masking-task group did. The matching-task group showed poorer conditioning, although in both groups conditioning occurred. The last study demonstrates the effectiveness of the masking task in manipulating awareness, but it also suggests the limitations of the procedure in eliminating awareness.

LEVELS OF CONDITIONING

A significant issue in conditioning theory has been concerned with the possibility that learning may occur at more than one level, and that perhaps behavior change may be going on at more than one level at the same time. Such a possibility had been suggested by Razran (1955) when he spoke of conditioning with perception, for example, perception of stimuli, and perception of relations between stimuli.

Later, Razran (1961) made clear that he felt that primitive forms of conditioning may occur without awareness and may serve as a basis for some unconscious process. He extended these views (1965, 1971) to what he termed evolutionary levels of learning. The distinctions made here bear resemblance to the difference proposed by Pavlov between first and second signal system conditioning, the first level involving direct sense impressions and the second involving symbolic and meaningful stimuli. It was implicit in Pavlov's theorizing that learning with the second signal system may take precedence over, inhibit, or at least interact with learning of the first signal system.

One of the arguments favoring a separation of types of learning was made by Bridger and Mandel (1964). They measured the electrodermal response of two groups of subjects. Both groups wore electrodes for receiving an electric shock and both were told that a shock would be delivered after a particular signal stimulus. Then, even though one group was given the shock and the other had only the threat of shock, both groups gave electrodermal responses to the signal stimulus that were greater than those given to the control stimulus. This was viewed as evidence for conditioning. Then both groups were told that there would be no more shocks and another series of signal trails was administered. The fact that the threat group extinguished and the shocked group did not was taken as evidence that different types of learning had occurred—one due to verbal factors only and the other due to actual experience with shock.

Extending upon this work and again using electrodermal response, Bridger and Mandel (1965) reported a study in which instructions to the effect that no more shock would occur (accompanied by removal of the shock electrodes) led to the abolition of a partial reinforcement extinction effect as compared to the retention of the effect in a noninformed group. The presence of a conditioned electrodermal response during extinction in the informed group combined with the abolition of the partial reinforcement effect was interpreted to indicate a two-component conditioned response characterized as a simple conditioned response and a mediated conditioned response.

In a more recent study by Mandel and Bridger (1967), six groups of subjects were studied, two with a 5-second interstimulus interval, two with a ½-second interstimulus interval, and two with a backward interstimulus interval. All groups were informed that the experiment contained three parts:

(*a*) a period when responses to a light alone would be recorded (preconditioning trials to CS); (*b*) a period when the light would be associated with electric shock; and finally, (*c*) a period during which responses to the light alone would again be recorded. All subjects were informed of the nature of the stimulus relations. Then, at the beginning of the third, or extinction, period one of the two groups in each of the interstimulus interval categories was interrupted by the experimenter removing the UCS electrodes, wiping the subject's skin clear of electrode jelly and reassuring him that no further shock would be given. This treatment of the informed subjects, as they were called, was contrasted with that of the other group in each condition where the experiment continued without interruption. A postexperimental inquiry was conducted to eliminate those subjects in the informed group who did not believe the extinction instructions.

The important extinction results reported were that among the two forward interstimulus interval groups, the 5-second group demonstrated a larger differential response during extinction under the noninformed condition whereas the ½-second groups had a larger differential response under the informed condition. The backward-conditioned response did not demonstrate resistance to extinction under either condition. It was concluded that both forward interstimulus interval groups demonstrated differential extinction responses under the informed condition which were contrary to their cognitive expectancies. This alleged maintenance of a differential response contrary to a cognitive expectancy was "considered be evidence for a level of GSR conditioning analogous to Mowrer's (1938) concept of true learning [Mandell & Bridger, 1967, p. 42]."

One of the dangers in drawing the above conclusion from the procedure used is similar to the difficulty encountered in trying to prove that true conditioning does not occur in the absence of awareness. In the latter (awareness) studies, inferences were made from a lack of evidence for conditioning, and it is argued that the other necessary conditions for true learning to occur have been present. In the Mandel and Bridger case, inferences were made from a lack of evidence for extinction, and it becomes necessary to assume that the subject does not have cognitive expectation. With human subjects this form of assumption is difficult to substantiate. In the study reported, the experimenters may be asking their subjects to report that the experimenter told them a falsehood. That is, they might have really expected a shock but have answered according to the demand characteristics of the situation.

It should also be pointed out that studies in which instruction variation is presumed to be the primary independent variable are often subject to inadvertent confounding of other types. An example in the present instance might be disinhibition (or dishabituation). It will be noted that the informed groups were interrupted for a period of time whereas the uninformed subjects had no such interruption. Disinhibition or changes in attitude toward the experiment might readily occur due to interruption.

The important question, however, remains whether this form of experimental demonstration indicates the existence of two types of learning, one based on cognitive expectancies and the other a noncognitive conditioning. It is concluded here that the evidence at hand is equivocal. It can be explained as resulting entirely from learning, developed through cognitive expectations, which is modified to varying degrees during extinction performance (rather than involving cognitive versus noncognitive processes). The data do argue strongly, however, that cognitive variables are important in determining extinction performance.

Wilson (1968) suggested an alternative to the "no more shock" extinction instructions for evaluating the relative contributions of simple and cognitive learning. Using a discrimination paradigm and a 1-second interstimulus interval, he instructed the subjects prior to acquisition and prior to extinction. At the beginning of extinction he told them that a discrimination reversal would occur (i.e., CS− would now be shocked and CS+ would not be shocked). He then administered stimuli without shock and found "a sudden reversal of the differential conditioning, GSRs of greater magnitude being evoked by

CS− [p. 491]." He concluded that the "study finds no evidence to support the contention that some part of a conditioned autonomic response is 'simple' in the sense that it is not mediated by *S*'s perception of stimulus contingencies [p. 493]." Subsequent study of the reversal procedure (Grings, 1973) has shown a clear dominance of cognitive variables for both long and short interstimulus intervals, and when reversal learning groups (negative transfer experience) are compared to groups whose previous training was pseudoconditioning (or nontransfer experience).

Another interpretation about levels was made by Fuhrer and Baer (1969) in the previously discussed study using a masking task. They argued that cognitive factors manifest themselves differently in the various components (topography) of the electrodermal response. They found that in their masking-task group both aware and unaware subjects show significant discrimination of first response (responses occurring just after the CS) but not significant second responses (occurring just before the onset of UCS). They conclude that they have demonstrated unaware conditioning.

INDIRECT APPROACHES TO "AWARENESS" VARIABLES

Most of the evidence considered thus far has relied upon manipulations of environmental conditions which are presumed to influence the subject's cognitive awareness or perception of the experimental situation. Such studies have been most numerous in electrodermal response conditioning. There is, however, considerable information from other types of studies—among them observations of special populations of verbal-conceptually handicapped subjects, studies using conditioned stimuli administered interoceptively, and research using the electrodermal response as an index of expectation or alertness.

In early efforts to evaluate the role of verbal factors in electrodermal response conditioning the writer used two special populations of subjects who were deficient in verbal capabilities. One of these was composed of totally deaf preschool-age children who, because of their deafness, were not equipped with verbal communication skills and who, because of age and deafness, would be expected to be very unsophisticated about conditioning (Grings, Lowell, & Honnard, 1961). These children showed conditioning as rapidly and reliably as college-age normal subjects.

The second investigation was made of adolescents confined in a state hospital for the mentally deficient. Two samples were chosen. One had an average IQ of 63 and possessed minimum verbal skills for ordinary communication. The other had a mean IQ of 34 and was almost completely nonverbal. It was anticipated that a difference in rate of acquisition of an autonomic conditioned discrimination would result from a difference in verbal ability. This did not occur, for both groups conditioned rapidly and reliably (Grings, Lockhart, & Dameron, 1962). Similar results have been reported by Baumeister, Beedle, & Urquhart (1964), whereas Lobb and Nugent (1966) and Lobb (1968) report an interaction of intelligence and interstimulus interval.

Still another avenue into the domain of cognitive–autonomic relations exists through the distinction between interoceptive and exteroceptive conditioning. Most human studies employ as conditioned stimuli exteroceptive stimuli like lights and tones. Few have used stimuli delivered interoceptively as has been done extensively with lower animals (e.g., Bykov, 1957). Stimuli to the interior of the body have a unique property in terms of awareness learning, for such stimuli are typically reported as only vague and diffuse in conscious experience.

Uno (1970) conducted a study comparing exteroceptive and interoceptive conditioning of the electrodermal response and made extensive observations of the role of perception and awareness. The interoceptive stimuli were different temperatures of a water-filled balloon, swallowed by the subjects and located midway in the esophagus. Exteroceptive CSs were tones of different frequencies, and the UCS was an electric shock. Differential conditioning was obtained to both interoceptive and exteroceptive stimuli. Analysis of subgroups in terms of their awareness of stimulus (CS) differences or of CS–UCS relations showed that unawareness of interoceptive stimuli or their reinforcement contingencies tended to eliminate the first response (the conditioned response occurring in the delay interval) but did not eliminate conditioning of the UCS omission response (the conditioned response occurring on trials where the UCS is omitted). Only one response was measured in the delay period so that a direct comparison cannot be made with the conclusions of Fuhrer and Baer (1969) about awareness and response topography.

There are other forms of research paradigms which have much in common with previously discussed manipulations of cognitive awareness, but which are typically not considered in connection with the above issues, possibly because the subject is assumed to be aware. Some of these will be noted briefly.

One is the motor avoidance task in which a CS serves as the signal for response. It is possible to observe conditioned anticipatory electrodermal responses if the time between CS and receipt (or nonoccurrence) of noxious stimulus is sufficiently long. Conclusions about awareness can be made from the correctness of the avoidance response. As an example of this type, Grings and Lockhart (1966) had subjects operate a four-position switch which was signalled by four visual stimuli: one always followed by shock, one never shocked, and two associated with switch positions which avoided the shock. Successive subgroupings exhibited progressively better avoidance learning with increased awareness instruction and showed significant decreases in magnitude of anticipatory electrodermal responses at the point of learning the avoidance contingency.

A related study (Colgan, 1970) produced similar results by making the receipt of shock predictable to the subject by instructing him about stimulation changes. With three distinctive stimuli, each was followed equally often by shock during an initial period of trials. Then half of the subjects were instructed that shock would follow only one stimulus and the other half were instructed that they would have a second series like the first. The preshock electrodermal response for the instructed groups stayed at a high magnitude for the cue still followed by shock and dropped sharply for the other two signals. For the other subgroup no between-stimulus differences occurred.

Any situation in which there is a distinctive signal for action to occur after a delay will provide electrodermal responses similar to those in more conventional delay conditioning. Two outstanding examples are the method of Ivanov-Smolensky where the signal to execute an instructed command takes the place of the conventional UCS (an example with electrodermal response conditioning would be that of Skorunskaia, 1958); and the reaction time task where the warning signal for the reaction time task can be made to have the properties of a discriminative CS (e.g., Baer & Fuhrer, 1969; Martin, 1965).

Related Theoretical Issues

The discussion to the present point permits the conclusion that in electrodermal conditioning situations many cognitive variables are of theoretical importance. The variables which have been identified stress the importance of conscious awareness and of verbally induced sets upon performance. It is clear that verbal expectancies develop in paired stimulation (conditioning) situations, and there is strong evidence that electrodermal response conditioning changes accompany these verbal awarenesses. Whether any learning of electrodermal behavior occurs in the absence of awareness is an open question. Of major importance, is the question of just how the cognitive factors operate in determining autonomic behavior change.

At the present time there is no comprehensive miniature theory to account for cognitive learning of autonomic behavior. There are, however, strong links to points of view that emphasize sign-significate learning (e.g., Tolman, 1932) and sequence-learning interpretations of conditioning (e.g., Grings, 1963; Woodworth, 1947) and with learned orienting and alerting behavior to stimuli (e.g., Sokolov, 1963).

What is central to all of these points of view is that the conditioned stimulus serves as a signal which informs the subject of the appearance of the second stimulus (the UCS). Some discussions about the nature of electrodermal conditioning (e.g., Grings, 1960, 1969; Grings & Sukoneck, 1971) emphasize concepts of perception, set, and expectancy based on temporal contingencies operating in the individual's environment. The autonomic responses are seen as arousal correlates to the preparation and expectation process. This suggests the organization of concepts underlining such learning in terms of perceptual variables of at least two main types: perception of stimuli and perception of relations or contingencies.

Paired stimulation experiences are assumed to change perceptions of the stimuli involved. A signal stimulus may take on the properties of the stimulus which it signals. Verbalizable properties of a conditioned stimulus appear to show alteration due to association with the UCS. This may occur with nonsymbolic as well as with symbolic stimuli. As an example of such change in nonsymbolic stimuli, one might note the work of Beck and Brooks (1967) and Sutterer and Beck (1970) who found semantic differential rating changes for neutral stimuli preceding and following a noxious stimulus. Generally speaking, the signal which is antecedent to an unpleasant event changes its affective quality toward unpleasantness, whereas the opposite is true for a stimulus following such an event (e.g., the change is toward pleasantness).

Shift in meaning of a symbolic stimulus implying a contingent event can be seen to change its arousal properties differently as a result of general exposure to the event, or graded experiences with the event (Grings & Zeiner, 1965). This type of phenomenon can be demonstrated to occur with symbolic stimuli not themselves reinforced but included in a group of reinforced stimuli (Grings, Carlin, & Appley, 1962). In other cases the intrinsic meaning of symbolic stimuli may determine the subject's perception of the stimulus sequence (e.g., Grant, 1972).

There is also support for the assumption that the perceptual properties of a noxious stimulus (e.g., the UCS) will change as a result of paired stimulation with a signal stimulus. For example, it has been generally reported that an electrodermal response to a noxious stimulus, like an electric shock, is less in magnitude when that stimulus is preceded by a signal than it is when it is presented alone. The phenomenon, termed UCR diminution (e.g., Kimmel, 1966) is complicated by difficulties arising from peripheral response interference (e.g., Grings & Schell, 1969) and orienting behavior (Badia & Defran, 1970).

In at least one study the rated intensity of the UCS was found to change in the same direction as the autonomic response (i.e., as the electrodermal response became less, the stimulus was rated less intense) (Schell & Grings, 1971). Other studies show different results (e.g., Kimmel, 1967; Furedy & Doob, 1971). One basis for predicting enhancement rather than diminution has been suggested in the situation where signal stimuli are followed by a UCS different from that with which they had previously been associated (Grings, 1960).

A theory for explaining both directions of response change (diminution and enhancement) in terms of perceptual changes has been proposed by Lykken (1962, 1968). His central concept is termed "preception" and involves a hypothetical mechanism for amplifying or attenuating the subjective intensity of predictable stimuli. Other evidence for the effect of stimulus pairings upon perception of the signaled stimulus comes from studies of preference for signaled and unsignaled shock (e.g., Badia, McBane, Suter, & Lewis, 1966); and judgments of stimulus intensity under different circumstances of warning (e.g., Epstein & Clark, 1970).

A central focus of research on awareness variables was upon perception of relations among stimuli or of contingencies. One immediate question concerns the manner in which the relation is learned, that is, how the subject becomes aware of a contingency. Most research to date falls into one of two groups. In one, learning is by some self-discovery method and in the other learning is by some form of verbal guidance. Common observation indicates that knowledge of relations or contingencies alone may be a weak modifier of behavior. Additional research concepts are needed which relate to the strength of belief about the knowledge of the contingency. One way such a variable can be studied is through reports of certainty by the subject. Other ways may be developed from considerations of the properties of the contingency relations.

One important feature of the relation in the case of simple sequence association is the information that is contained in the sequence. Two major kinds of information occur. One concerns the qualities or characteristics of the event to follow and the other arises from the certainty elements of the relation. In the last category at least three forms of certainty are available for research: event certainty, time certainty, and quality certainty. Examples of event certainty can be developed through manipulation of the proportion of times that the second event follows the first (a concept somewhat analogous to percentage of reinforcement, except that it might include pattern as well as rate of occurrence of pairs, etc.). Time certainty can be defined by the difference between a constant interstimulus interval and a variable interstimulus interval, for the arrival of the unpleasant signaled event will be more uncertain with variable interstimulus interval than with constant interstimulus interval. Quality uncertainty relates to the extent to which the qualitative properties of the second stimulus are predictable, and is most simply defined by contrasting constant intensity and variable intensity second stimuli. Recent experiments to evaluate these informational variables in the present context manipulated event certainty (eg., Epstein & Roupenian, 1970; Grings & Sukoneck, 1971), time certainty (e.g., Grings & Schell, 1971), and quality certainty (e.g., Epstein & Clarke, 1970; Ohman, 1971).

In brief, it is concluded that cognitive determination of autonomic conditioning results from the central fact that the CS contains information properties which can be conceptualized in terms of knowledge of stimulus relations and knowledge of stimulus properties. Experimental manipulation can be achieved through concepts of certainty, subjective probability, and information transmission. There is an implied role played by the signal. In aversive situations this is postulated as controlling (modulating) the arousal characteristics of the strong (unconditioned) stimulus. Finally, there is an implication that the response to the unconditioned stimulus is dependent to a degree upon the information provided by the signal.

REFERENCES

Badia, P., McBane, B., Suter, S., & Lewis, P. Preference behavior in an immediate versus variably delayed shock situation with and without a warning signal. *Journal of Experimental Psychology,* 1966, **72**, 847–852.

Badia, P., & Defran, R. H. Orienting responses and GSR conditioning: A dilemma. *Psychological Review,* 1970, **77**, 171–180.

Baer, P. E., & Fuhrer, M. J. Cognitive factors in differential conditioning of the GSR: Use of a reaction time task as the UCS with normals and schizophrenics. *Journal of Abnormal Psychology,* 1969, **74**, 544–552.

Baumeister, A. A., Beedle, R., & Urquhart, D. GSR conditioning in normals and retardates. *American Journal of Mental Deficiency,* 1964, **69**, 114–120.

Beck, R. C., & Brooks, C. I. Human judgments of stimuli associated with shock onset and termination. *Psychonomic Science,* 1967, **8**, 327–328.

Bridger, W. H., & Mandel, I. J. A comparison of GSR fear responses produced by threat and electric shock. *Journal of Psychiatric Research,* 1964, **2**, 31–40.

Bridger, W. H., & Mandel, I. J. Abolition of the PRE by instructions in GSR conditioning. *Journal of Experimental Psychology,* 1965, **69**, 476–482.

Bykov, K. M. *The cerebral cortex and the internal organs.* (Trans. by W. H. Gantt) New York: Chemical Publishing Co., 1957.

Chatterjee, B. B., & Eriksen, C. W. Conditioning and generalization of GSR as a function of awareness. *Journal of Abnormal and Social Psychology,* 1960, **60**, 396–403.

Chatterjee, B. B., & Eriksen, C. W. Cognitive factors in heart rate conditioning. *Journal of Experimental Psychology,* 1962, **64**, 272–279.

Colgan, D. M. Effects of instructions upon the skin resistance response. *Journal of Experimental Psychology,* 1970, **86**, 108–112.

Cook, S. W., & Harris, R. E. The verbal conditioning of the galvanic skin reflex. *Journal of Experimental Psychology,* 1937, **21**, 202–210.

Dawson, M. E. Cognition and conditioning: Effects of masking the CS-UCS contingency on human GSR classical conditioning. *Journal of Experimental Psychology,* 1970, **85**, 389–396.

Dawson, M. E., & Grings, W. W. Comparison of classical conditioning and relational learning. *Journal of Experimental Psychology,* 1968, **76**, 227–231.

Dawson, M. E., & Reardon, P. Effects of facilitory and inhibitory sets on GSR conditioning and extinction. *Journal of Experimental Psychology,* 1969, **82**, 462–466.

DeNike, L. D. The temporal relationship between awareness and performance in verbal conditioning. *Journal of Experimental Psychology,* 1964, **68**, 521–529.

Diven, K. Certain determinants in the conditioning of anxiety reactions. *Journal of Psychology,* 1937, **3**, 291–308.

Epstein, S., & Clarke, S. Heart rate and skin conductance during experimentally induced anxiety: Effects of anticipated intensity of noxious stimulation and experience. *Journal of Experimental Psychology,* 1970, **84**, 105–112.

Epstein, S., & Roupenian, A. Heart rate and skin conductance during experimentally induced anxiety: The effect of uncertainty about receiving a noxious stimulus. *Journal of Personality and Social Psychology,* 1970, **16**, 20–28.

Fuhrer, M. J., & Baer, P. E. Differential classical conditioning: Verbalization of stimulus contingencies. *Science,* 1965, **150**, 1479–1481.

Fuhrer, M. J., & Baer, P. E. Cognitive processes in differential GSR conditioning: Effects of a masking task. *American Journal of Psychology,* 1969, **82**, 168–180.

Furedy, J. J., & Doob, A. N. Autonomic responses and verbal reports in further tests of the preparatory-adaptive-response interpretation of reinforcement. *Journal of Experimental Psychology,* 1971, **89**, 258–264.

Grant, D. A. Consequences of processing and a preliminary model for processing information conveyed by verbal conditioned stimuli. In A. Black & W. Prokasy (Eds.), *Classical Conditioning II.* New York: Appleton-Century-Crofts, 1972.

Grings, W. W. Preparatory set variables related to classical conditioning of autonomic responses. *Psychological Review,* 1960, **67**, 243–252.

Grings, W. W. Classical conditioning. In M. Marx (Ed.), *Theories in contemporary psychology.* New York: Macmillan, 1963.

Grings, W. W. Verbal-perceptual factors in the conditioning of autonomic responses. In W. F. Prokasy (Ed.), *Classical conditioning.* New York: Appleton-Century-Crofts, 1965.

Grings, W. W. Anticipatory and preparatory electrodermal behavior in paired stimulation situations. *Psychophysiology,* 1969, **5**, 597–611.

Grings, W. The role of consciousness and cognition in autonomic behavior change. In F. J. McGuigan & R. Schoonover (Eds.), *The psychophysiology of thinking.* New York: Academic Press, 1973, in press.

Grings, W., Lowell, E., & Honnard, R. GSR conditioning with preschool age deaf children. *Journal of Comparative and Physiological Psychology,* 1961, **54**, 143–148.

Grings, W., Lockhart, R., & Dameron, L. Conditioning autonomic responses of mentally subnormal individuals. *Psychological Monograph,* 1962, **76**(39, Whole No. 558).

Grings, W., Carlin, S., & Appley, M. Set, suggestion and conditioning. *Journal of Experimental Psychology,* 1962, **63**, 417–442.

Grings, W., & Lockhart, R. Effects of "anxiety lessening" instructions and differential set development on the extinction of the GSR. *Journal of Experimental Psychology,* 1963, **66**, 292–299.

Grings, W., & Lockhart, R. Galvanic skin response during avoidance learning. *Psychophysiology,* 1966, **3**, 29–34.

Grings, W. W., & Schell, A. M. Magnitude of electrodermal response to a standard stimulus as a function of intensity and proximity of a prior stimulus. *Journal of Comparative and Physiological Psychology,* 1969, **67**, 77–82.

Grings, W., & Schell, A. The effects of trace versus delay conditioning, ISI variability, and instructions on UCR diminution. *Journal of Experimental Psychology,* 1971, **90**, 136–140.

Grings, W., & Sukoneck, H. Prediction probability as a determiner of anticipatory and preparatory behavior. *Journal of Experimental Psychology,* 1971, **91**, 310–317.

Grings, W., & Zeiner, A. Autonomic responses to words modified by sensitizing and conditioning experiences. *Journal of Psychosomatic Research,* 1965, **8**, 373–378.

Haggard, E. A. Experimental studies in affective processes: I. Some effects of cognitive structure and active participation on certain autonomic reactions during and following experimentally induced stress. *Journal of Experimental Psychology,* 1943, **33**, 257–284.

Harvey, B., & Wickens, D. D. Effect of instructions on responsiveness to the CS and to the UCS in GSR conditioning. *Journal of Experimental Psychology,* 1971, **87**, 137–140.

Hilgard, E. R., & Humphreys, L. G. The effect of supporting and antagonistic voluntary instructions on conditioned discrimination. *Journal of Experimental Psychology,* 1938, **22**, 291–304.

Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century-Crofts, 1940.

Hill, F. A. Effects of instructions and subject's need for approval on the conditioned galvanic skin response. *Journal of Experimental Psychology,* 1967, **73**, 461–467.

Kimmel, E. Judgments of UCS intensity and diminution of the UCR in classical GSR conditioning. *Journal of Experimental Psychology,* 1967, **73**, 532–543.

Kimmel, H. D. Inhibition of the unconditioned response in classical conditioning. *Psychological Review,* 1966, **73**, 232–240.

Lacey, J. I., Smith, R. L., & Green, A. Use of conditioned autonomic response in the study of anxiety. *Psychosomatic Medicine,* 1955, **17**, 208–217.

Lobb, H. Trace GSR conditioning with benzedrine in mentally defective and normal adults. *American Journal of Mental Deficiency,* 1968, **73**, 239–246.

Lobb, H., & Nugent, C. M. Interaction between intelligence level and interstimulus trace interval in electrodermal conditioning. *American Journal of Mental Deficiency,* 1966, **70**, 548–555.

Lykken, D. T. Preception in the rat: Autonomic response to shock as function of length of warning interval. *Science,* 1962, **136**, 665–666.

Lykken, D. T. Neuropsychology and psychophysiology in personality research. In Borgatta & Lambert (Eds.), *Handbook of personality theory and research*. New York: Rand McNally, 1968.

Mandel, I. J., & Bridger, W. H. Interaction between instructions and ISI in conditioning and extinction of the GSR. *Journal of Experimental Psychology,* 1967, **74**, 36–43.

Martin, I. Discriminatory GSRs. *Activitas Nervosa Superior,* 1965, **7**, 217–223.

Mowrer, O. H. Preparatory set. *Psychological Review,* 1938, **45**, 62–91.

Norris, E. B., & Grant, D. A. Eyelid conditioning as affected by verbally induced inhibitory set and counter reinforcement. *American Journal of Psychology,* 1948, **61**, 37–49.

Ohman, A. Interaction between instruction-induced expectancy and strength of unconditioned stimulus in GSR conditioning. *Journal of Experimental Psychology,* 1971, **88**, 384–390.

Razran, G. Conditioning and perception. *Psychological Review,* 1955, **62**, 83–95.

Razran, G. The observable unconscious and inferable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychological Review,* 1961, **68**, 81–147.

Razran, *G*. Evolutionary psychology: Levels of learning and perception and thinking. In B. Wolman (Ed.), *Scientific psychology*. New York: Basic Books, 1965.

Razran, G. H. *Mind in evolution*. Boston: Houghton Mifflin, 1971.

Schell, A. M., & Grings, W. W. Judgments of UCS intensity and diminution of the unconditioned GSR. *Psychophysiology,* 1971, **8**, 427–432.

Shean, G. D. The relationship between ability to verbalize stimulus contingencies and GSR conditioning. *Journal of Psychosomatic Research,* 1968, **12**, 245–249.

Silverman, R. E. Eliminating a conditioned GSR by the reduction of experimental anxiety. *Journal of Experimental Psychology,* 1960, **59**, 122–125.

Skorunskaia, T. N. Interaction of signal systems during simultaneous presentation of direct and verbal stimuli of opposite significance. *Pavlov Journal of Higher Nervous Activity,* 1958, **8**, 327–332.

Sokolov, Y. N. *Perception and the conditioned reflex*. New York: Macmillan, 1963.

Sutterer, J., & Beck, R. Human response to stimuli associated with shock onset and termination. *Journal of Experimental Psychology,* 1970, **4**, 163–170.

Tolman, E. C. *Purposive behavior in animals and men*. New York: Appleton-Century-Crofts, 1932.

Uno, T. The effects of awareness and successive inhibition on interoceptive and exteroceptive conditioning of the galvanic skin response. *Psychophysiology,* 1970, **7**, 27–43.

Wickens, D. D., Allen, C. K., & Hill, F. A. Effects of instruction on extinction of the conditioned GSR. *Journal of Experimental Psychology,* 1963, **66**, 235–240.

Wilson, G. D. Reversal of differential GSR conditioning by instructions. *Journal of Experimental Psychology,* 1968, **76**, 491–493.

Woodworth, R. S. Reinforcement of perception. *American Journal of Psychology,* 1947, **60**, 119–124.

(Received January 11, 1972)

Psychological Bulletin
1973, Vol. 79, No. 3, 211–214

COMMENT ON "LINGUISTIC DEFICIENCY AND THINKING: RESEARCH WITH DEAF SUBJECTS 1964–1969"[1]

HARRY BORNSTEIN AND HOWARD L. ROY[2]

Gallaudet College

A paper by Furth reviewed 39 studies in which the cognitive performances of deaf children were compared with those of hearing children. With some exceptions, deaf children performed as well as hearing children on the tasks set for them. Because Furth believes that deaf subjects are linguistically deficient, he concluded that the thinking processes of deaf children are similar to those of hearing children and therefore must be explained without recourse to verbal processes. He offers as evidence for language deficiency one standardization study which shows that the mean reading achievement of deaf students on a nationally standardized achievement test falls below fourth-grade equivalence. This paper argues (*a*) that the achievement test was normative referenced and does not provide evidence for an inability to handle English sentences and (*b*) that there exists still other evidence of deaf student mean achievement at fourth- and fifth-grade equivalence. We contend that deaf subjects cannot be regarded as language deficient without explicit demonstration to that effect. Failing this, the reviewed studies cannot be said to contribute evidence for or against the hypothesis that language is not related to or required for cognitive development.

In a recent issue Furth (1971) reviewed a number of studies in which the cognitive performances of deaf children were compared with those of hearing children. The tasks involved in these studies were categorized as: rule learning, logical symbols, Piaget-type tasks, memory, and perception. With some exceptions, deaf children performed as well as hearing children on the tasks set for them. On this basis, Furth concluded "that thinking processes of deaf children are similar to those of hearing children and therefore must be explained without recourse to verbal processes [p. 70]."

The fundamental rationale for this conclusion rests on the assumption by Furth that deaf subjects are linguistically deficient. Thus, any cognitive task on which deaf persons do as well or better than hearing subjects can be viewed as a performance which does not require language. He devotes a considerable section of his article to defining and supporting this conclusion. Our comments are directed at that section, but it should be noted that we are not necessarily opposed to the point of view that cognitive processes or reasoning ability are not dependent upon language. We do object, however, to Furth's contention that deaf subjects are lacking in the language skills necessary or appropriate for the cognitive tasks in the reviewed studies. In our opinion, the data obtained from deaf subjects in these studies do not provide adequate tests of his thesis.

One of the first studies reviewed by Furth, a study by Blank and Bridges (1966), was concerned with learning in situations involving absence (negation or numbers). The relevant cue was the presence or absence of stimulation (light or vibration). The deaf children (3–6 years of age) did as well as the hearing children. The authors concluded that deaf children, although deficient in general language, succeeded "not because they had a non-language concept, but rather because they possessed the appropriate symbol (e.g., word, gesture) to designate the concept [p. 59]." Interestingly enough, Furth does not quarrel with this conclusion. It seems, therefore, that these deaf subjects are linguistically deficient only if one equates deficiency with an inability to "handle" sen-

---

[1] The authors are grateful to James M. Pickett, Horace N. Reynolds, William C. Stokoe, Jr., and Augustine Gentile for their suggestions and comments on earlier drafts of this paper.

[2] Requests for reprints should be sent to Harry Bornstein, Department of Psychology, Gallaudet College, Kendall Green, Washington, D. C. 20002.

TABLE 1

MEAN GRADE EQUIVALENTS BY TEST BATTERY LEVEL FOR THE PARAGRAPH MEANING SUBTEST OF THE STANFORD ACHIEVEMENT TEST FOR HEARING IMPAIRED STUDENTS WITH BETTER EAR LOSS OF 60 AND ABOVE

Age	Intermediate I level		Intermediate II level		Advanced level		Weighted mean of means	
	n	Mean grade equivalent	*n*	Mean grade equivalent	*n*	Mean grade equivalent	*n*	Grade equivalents
12	159	3.35					159	3.35
13	222	3.39	166	4.08			388	3.68
14	174	3.20	179	4.43			353	3.82
15	191	3.24	192	4.31	192	5.54	575	4.36
16	151	3.26	200	4.30	258	5.77	609	4.66
17			148	4.10	257	5.90	405	5.24
18					235	5.79	235	5.79
19					159	5.68	159	5.68

Note. Adapted from A. Gentile & S. Di Francesca. *Academic achievement test performance of hearing impaired students in the United States: Spring 1969.* (Series D., No. 1) Washington, D. C.; Gallaudet College, Office of Demographic Studies, 1969.

tence structure. In fact, however, the deaf students in this study were not linguistically deficient in terms of the tasks required. They knew the word or gesture needed to perform the cognitive tasks presented to them. We think it is incumbent upon Furth to demonstrate for each such study that a word, phrase, or series of phrases is not sufficient language to perform the task required. Until this has been established, no conclusion about language deficiency relative to cognitive function can be drawn from using deaf students as subjects.

There is another matter which gives us some concern. It is Furth's definition of language and his evidence in support of it. He states that ". . . comprehension of fourth-grade reading as measured by present standardized tests can be suggested as an appropriate criterion [of language competence] . . . [p. 69]." He then offers results from a single standardization study based upon a sample of 5,284 deaf pupils between the ages of 10½ and 16½ conducted in 1963. The test was the Metropolitan Elementary Reading Test. The mean grade equivalent of the oldest group was 3.6 and this represented very little change from the performance of the younger groups. Furth considers these findings as conclusive evidence of language deficiency. Two points can be noted: First, the Metropolitan (as well as the Stanford Achievement Test which will be cited later) is a normative and not a criterion referenced test. Test scores expressed as grade equivalences represent a performance which is "equivalent" to some modal score achieved by children at a given grade—in this case 3.6. It is very difficult to describe what this level of skill really represents, but if the reader doubts that materials included in a third-grade reader are made up of sentences, he need merely glance at such a book. Further, test constructors normally do not interpret a grade equivalence score of 3.6 to mean that the group cannot read some fourth-, fifth-, or sixth-grade material. Normative referenced tests are simply not designed to furnish such conclusions. Interested readers may refer to Angoff (1971) for a current scholarly review of permissible interpretations of age and grade equivalence scores. The second point is that more recent evidence on the reading abilities of deaf students differed from the earlier Metropolitan Test findings. Gentile and DiFrancesca (1969) conducted the first annual survey of Stanford Achievement Test performance of hearing impaired students in the United States. Test results for 11,544 pupils from ages 7 to 19 were reported. Table 1 is based on data from that report and shows mean grade equivalents by age for the Paragraph Meaning subtest of the Stanford Achievement Test. This subtest is included at several different test battery levels and is administered by the school at a level thought appropriate for any given child. The data in Table 1 deal only with those students with hearing losses of 60 decibels or more in the better ear. The data do not show an inspiring reading performance. However, two-thirds of the age means are at the fourth and higher grade levels. Fifth grade appears to be the mean achievement ceiling. We are not prepared to say that these data are more valid than those secured earlier with another test, but surely the evidence does not point uneqivocally to reading achievement below fourth-grade equivalence, whatever that may mean in absolute terms.

It seems to us that Furth confuses inadequate performance or limited competence in handling the grammatical structure of English with deficiency or inability to form meaningful English sentences. We believe that it is scientifically more useful and more consistent with current linguistic theory to view the general structure of English as made up of a very complex set of phrase structure and transformational rules. Deaf persons differ from hearing persons in their ability to use these rules. There is a considerable body of direct information available on which rules and in what settings the deaf have difficulty. There is, however, no direct empirical information available on a claimed inability of deaf persons to handle the syntactical and grammatical structure of the English language, possibly because it is a global concept which has not been operationally defined.

Any discussion of the language of deaf persons is further complicated by the fact that many or most of them have a second language system available to them, that is, the American Sign Language. Furth (1971) states:

> The conventional gestural system sign language in use among deaf students is only available in childhood to those deaf children who are born of deaf parents—a very small percentage of the deaf population. Hence it is literally true that most deaf children grow up during the early formative period of life without any systematized conventional language. The significant fact is that in spite of not having been provided with a conventional symbol structure, deaf children construct their own symbols as they are needed for the development of thinking. On this crucial fact rests the present writer's interpretation of the reviewed studies [p. 69].

TABLE 2

BACKGROUND AND SIGN LANGUAGE FLUENCY INFORMATION ON 1968–1969 GALLAUDET ENTERING STUDENTS ($N = 220$)

Background Information	n	%
Sign language ability		
Fluent	114	52
Fairly well	71	32
A little	18	8
Not at all	14	6
No response	3	2
Kind of school attended		
Public hearing school	58	26
Day school for deaf	35	16
Residential school for deaf	135	61
No response	8	4
Total	236	107
Age sign language learned		
0–5 years	51	23
6–10	79	36
11–15	38	17
16–20	28	13
Over 20	3	1
No response	21	10
From whom learned		
Parents	29	13
Staff	81	37
Friends	139	63
Other	13	6
No response	9	4
Total	271	133[a]

Note. Reprinted from H. Bornstein & B. Kannapell. New signs for instructional purposes. (Final Report, OE Project No. 6 = 1924) Washington, D. C.: United States Office of Education, 1969.

[a] Some students attended more than one kind of school or learned the sign language from more than one kind of person.

There are several observations which can be made about this statement. First, no formal evidence is offered to support it. Second, while no age numbers are provided for "the early formative period of life," the really crucial consideration is the age of the students involved in the reviewed studies. Put in another way, is it not possible that the students studied were provided with a usable conventional symbol structure by others? An analysis of the upper and lower extremes of the range of ages of the children serving as subjects for the reviewed studies indicates that only the youngest children in eight studies were below 6 years of age, the normal beginning age for school attendance. The very youngest children in the remaining 31 studies are school-age students. There is at least one survey by Bornstein and Kannapell (1969) which has attempted to determine when and from whom deaf students learned the sign language. It is summarized in Table 2. The students were the entering 1968–1969 class at Gallaudet College, not by any means a representative sample of deaf pupils. However, the data in Table 2 do reveal some interesting information about how and when the sign language is learned in the United States. Most of the group report that they learned the sign language *after* reaching the school setting from friends and school staff.

Since school age corresponds to the age of most of the subjects participating in the studies reviewed by Furth, it is very possible that this linguistic system was indeed available to some or many of the subjects. Just how many would know the sign language is difficult to determine, but it is certainly Furth's burden to demonstrate that the subjects did not have this linguistic system at their disposal if he concludes that language could not have mediated the cognitive tasks involved in the studies. Lest the reader unfamiliar with the American Sign Language assume that this is a very primitive language system, it can be stated that the American deaf have a vocabulary of from 3,000 to 5,000 signs at their disposal and they use signs in a syntax peculiar to that language (Stokoe, Casterline, & Croneberg, 1965). The description of the studies in Furth's article leads us to believe that many cognitive performances described might well be expressed or mediated in the sign language.

REFERENCES

ANGOFF, W. H. Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), *Educational Measurement.* (2nd ed.) Washington, D. C.: American Council on Education, 1971.

BLANK, M., & BRIDGES, W. H. Conceptual cross-modal transfer in hearing and deaf children. *Journal of Child Experimental Psychology,* 1969, 7, 407–408.

BORNSTEIN, H., & KANNAPELL, B. New signs for instructional purposes. *ERIC*, 1969, No. ED-042-293.

FURTH, H. Linguistic deficiency and thinking: Research with deaf subjects 1964–69. *Psychological Bulletin,* 1971, 76, 58–72.

GENTILE, A., & DI FRANCESCA, S. *Academic achievement test performance of hearing impaired students in United States: Spring 1969.* (Series D, No. 1) Washington, D. C.: Gallaudet College, Office of Demographic Studies, 1969.

STOKOE, W., CASTERLINE, D., & CRONEBERG, C. *Dictionary American Sign Language.* Washington, D. C.: Gallaudet College Press, 1965.

(Received January 14, 1972)

Psychological Bulletin
1973, Vol. 79, No. 3, 215–216

FURTHER THOUGHTS ON THINKING AND LANGUAGE

HANS G. FURTH [1]

Catholic University of America

Deaf children are characterized as deficient, that is, limited, in their competence in handling linguistic rules, not necessarily in knowing simple words or signs. It is suggested that Bornstein and Roy's theoretical preconditions for studying the language-thinking hypothesis are experimentally untestable. An adequate interpretation of their tables on reading and language confirms rather than weakens the claim of limited linguistic competence in deaf children.

Bornstein and Roy (1973) claim that they do not necessarily oppose my conclusions (Furth, 1971) but they consider my description of linguistic deficiency inadequate. On a terminological level, they would prefer the term "limited competence." I chose the term "deficiency" in analogy with mental deficiency which implies some limitation in mental competence, not a total lack.

On the linguistic status of deaf persons the authors unfortunately give only scant clarifications so that this important topic remains to be studied. In the meantime anyone can observe—and this is not disputed by Bornstein and Roy—that the great majority of deaf children have a strikingly limited knowledge of any linguistic system. Moreover, the authors accept (and in part document) my claim that most deaf children learn sign language only after entering school. I have explicitly stated that I focus on the knowing or not knowing of a linguistic system and not of single words. I have indeed equated linguistic deficiency with a "limited competence in handling a grammatical structure."

One could hypothesize that for intellectual development the first 6 years are crucially important and the absence of a conventional symbol system such as English or sign language during these years could have harmful effects on the developing intelligence. It is of scientific interest to demonstrate that the absence of a systematized conventional language during the early formative years of life has few deleterious effects on intellectual development. This is what I have tried to show in Furth (1971) and have further documented, including social and personality aspects, in Furth (1973).

The main theoretical objection of the authors to my position is contained in such phrases as, "language skills necessary or appropriate for the cognitive task" or "language skills in English or sign language for any given cognitive task." I have never claimed and do not need to claim in support of my thesis that deaf children do not know the potentially correct word, gesture, or sign for enunciating the relevant principle of some tasks (e.g., "no," "same," "blue," "big," "three," etc.). The authors feel that "it is incumbent upon Furth to demonstrate, for each study, that a word, phrase, or series of phrases is not sufficient language to perform the task required [p. 212]." This statement may appear plausible and meaingful to someone who holds the theory that a linguistic medium is a prerequisite for intelligent performance. However, what experimental design could conceivably demonstrate that certain words or phrases are *not sufficient* for a successfully performed task? It is known from animal experiments that a simple- and double-alternation task can be performed *without* language. A 4-year-old child may know the words 'left, left, right, right" and nevertheless fail on double alternation. But how could one go about showing that for an 8-year-old child who succeeds on double alternation, knowing the phrase "left, left, right, right" is not sufficient? I wonder how the authors would interpret the performance I reported in Study 20 on probability, or in Studies 21 and 22 on visual transformation. Apart from the question of sufficiency, what language skills could be relevantly investigated?

---

[1] Requests for reprints should be sent to Hans G. Furth, Center for Research in Thinking and Language, Catholic University of America, Washington, D. C. 20017.

As to Table 1 in the Bornstein and Roy paper, which is adduced as evidence of fifth-grade achievement in reading for deaf youngsters, it can be seen readily that these scores are primarily a function of the scoring floor of the different tests. For the same tests no improvement at all takes place over an age range of 5 years; the only improvement occurs across different tests with scores that are rather close to chance levels for the respective tests (2.8, 3.6, 4.8).

Table 2 is based on a highly select sample. But it does confirm what I have reported. Only a small minority (13%) of the 220 students learned sign language from their parents. The additional finding that only 53% of this select sample is fluent in sign language is particularly illuminating as it indicates that even after leaving high school a great many deaf youngsters are at home neither in English nor in sign language. They are linguistically deficient in the precise sense of the word, in contrast to hearing persons who are all fluent in at least one language from 4 years of age onwards.

REFERENCES

Bornstein, H., & Roy, H. L. Comment on "Linguistic deficiency and thinking: Research with deaf subjects 1964–1969." *Psychological Bulletin*, 1973, **79**, 211–214.

Furth, H. G. Linguistic deficiency and thinking: Research with deaf subjects 1964–69. *Psychological Bulletin*, 1971, **76**, 58–72.

Furth, H. G. *Deafness and learning: A social-developmental psychology*. Belmont, Calif.: Wadsworth, 1973.

(Received October 6, 1972)

VOL. 79, NO. 4 APRIL 1973

Psychological Bulletin

© 1973 by the American Psychological Association, Inc.

MORAL CONDUCT AND MORAL CHARACTER:

A PSYCHOLOGICAL PERSPECTIVE[1]

ROBERT HOGAN[2]

Johns Hopkins University

This paper suggests that moral conduct can be explained, and the development of moral character can be described, in terms of five concepts or dimensions: moral knowledge, socialization, empathy, autonomy, and a dimension of moral judgment. These dimensions arise from a conception of man as a rule-formulating and rule-following animal, and a view of morality as a methodology, a result of biological and cultural evolution, which regulates and moderates social conduct. Each concept is given a theoretical and an operational definition, its developmental antecedents are discussed, and evidence is provided concerning its usefulness as a means for understanding moral conduct. Finally, the model as a whole is evaluated.

The subject of moral conduct contains some of the most intriguing problems and paradoxes in the social sciences. These problems (e.g., why a man would follow the path of greatest resistance; why saints seem obsessed with a sense of sin) have preoccupied social thinkers since Plato, including, more recently, Durkheim, Weber, Freud, Piaget, William James, and George Herbert Mead. In the early stages of American psychology there was great interest in the "psychology of moral conduct"; however, the Character Education Inquiry (Hartshorne & May, 1930) seemed to demonstrate that most such behavior is situation specific, and the subject fell into subsequent neglect. In a series of papers beginning in 1958, Kohlberg revived interest in moral psychology, and his research on moral judgment exemplified the relevance, complexity, and psychological richness traditionally associated with the topic. This paper describes a new psychological perspective on moral conduct, and presents some empirical findings which follow from this perspective. The paper begins by outlining the assumptions and definitions required for the subsequent discussion. The burden of the paper, however, is to argue that development of moral character can be described, and moral conduct can be explained in terms of five concepts or dimensions: moral knowledge, socialization, empathy, autonomy, and a dimension of moral judgment. Each dimension is defined both conceptually and operationally, its developmental antecedents are described, and evidence is presented concerning its empirical utility for understanding moral conduct and character development. Finally, the model taken as a whole is discussed.

ASSUMPTIONS

An earlier paper (Hogan & Henley, 1970) suggested that as an alternative to explaining social behavior in terms of needs, drives, traits, or mechanisms of coping and defense, it might be useful to consider man as a rule-formulating and rule-following animal. The

---

[1] Work on this paper was supported by the United States Office of Education Grant No. OEG-2-7-061610-0207 to the Johns Hopkins Center for the Social Organization of Schools. The author wishes to thank Ellen Dickstein, Catherine Garvey, Esther Greif, William Kurtines, and Roger Webb for their critical comments on earlier versions of this paper.

[2] Requests for reprints should be sent to Robert Hogan, Department of Psychology, Johns Hopkins University, Baltimore, Maryland 21218.

I 705 (5)

argument derives from a more comprehensive theory of social behavior based on psychological (as opposed to sociological) role theory, and an image of man as a game-playing animal (cf. Goffman, 1959; Huizinga, 1955; Peters, 1958; Wright, 1971). While the details of the argument are not important here, three of its assumptions are (the first two assumptions are substantive, the third follows as their methodological consequence). It is assumed first that all purposive social behavior occurs within matrices of overlapping human rule systems (e.g., grammars, legal codes, rules of courtesy) which have varying degrees of specificity. (It follows, somewhat tautologically, that behavior not guided by rules is nonpurposive or "random.") Second, associated with every human rule system is an ethic, variously called sportsmanship, fairness, justice, equity, or the spirit of the game. Huizinga (1950), for example, remarks that, "In every . . . community that is healthy . . . there will blossom an ideal of chivalry and chivalrous conduct, whether it be in Greece, Arabia, Japan, or medieval Christendom [p. 64]." This ethic, usually known and acknowledged by seasoned participants of a given rule system, is a spontaneous natural emergent, a product of role-taking experience and cognitive development. Its function is to provide a perspective from which the equity of conventional rules may be judged. Third, the broad outlines and general patterns of social behavior can be explained in terms of formal properties of various applicable rule systems. Individual differences, however, must be accounted for in terms of the differing fashions in which persons think about and use rules. That is, many of the similarities among members of the same family, ethnic group, or social class are attributable to structural variables (the rules which apply within the group), while the differences among these people are a function of how each person uses rules.

This paper also assumes that morality is a natural phenomenon, an adaptive response to evolutionary pressures, and that an understanding of moral behavior is relative to our knowledge of man's biological and psychological nature. A large number of biological mechanisms undoubtedly conspire to provide the preconditions for moral conduct; however, four are particularly relevant to the model presented below. First, as Waddington (1967) argues, the dynamics of human culture necessarily presuppose "the role of authority-acceptor." Consequently, a disposition to comply with authority is considered to be part of man's innate constitution, and this disposition is seen most vividly in children. Second, Darwin (1872) suggested in *The Expression of the Emotions in Man and Animals* that man is naturally disposed toward altruistic feelings of group loyalty. The biological nature of altruistic tendencies has been further discussed more recently by Campbell (1965) and Hamilton (1971); the implication of their arguments is that man is naturally sensitive to the expectations and demands of his peers, and this sensitivity leads to social compliance. Third, Lorenz (1966) points out that there is a strong tendency for rule-governed behavior to become ritualized, autonomous, and performed for its own sake. This suggests a biological basis for autonomous moral conduct. Finally, Tinbergen (1968), under the concept of *Sollwert* (should-value), describes the phenomenon in which the young of a species internalize certain adult behavior patterns during "sensitive periods," and these *Sollwerte* are used as internal templates against which young animals compare and evaluate their later performances. *Sollwerte*, then, seem to serve as innate mechanisms for self-regulation. The four biological concepts just described have obvious relevance to socialization, empathy, autonomy, and internal standards for the evaluation of behavior, the sociopsychological processes underlying the development of moral character and conduct.

Many contemporary psychologists consider moral action to be in some way distinct from ordinary social conduct. Thus Kohlberg[3] remarks that in the moral domain, persons feel that rules are constraining in and of themselves, whereas in the social (nonmoral) area, persons recognize the content of convention as artificial. Wright (1971) similarly observes, "Both the conceptual framework in which they are formulated and the experience which goes with them mark out moral rules as distinctively different from, for example, the rules of tennis, grammar, or etiquette [p. 14]." In contrast, the present paper makes no sharp distinction between social and moral rules. Rather, it assumes that an evaluative tendency is an integral part of social conduct, and that morality has a social job to do; that is, the function of morality is to regulate and moderate human affairs. The following quotations express the flavor of this assumption:

---

[3] L. Kohlberg. The development of modes of moral thinking and choice in the years 10 to 16. Unpublished doctoral dissertation, University of Chicago, 1958.

> Moral conduct is essentially social conduct, and there could be no serious objection to the use of the two expressions as synonymous; but it is more in conformity with common usage to restrict the term "moral" to the higher forms of social conduct of which man alone is capable [McDougall, 1908, p. 150].

> Manners in their best sense . . . cannot be distintinguished from morals, and are nothing but the expression of the courtesy of an individual toward people about him [Mead, 1934, p. 263].

These remarks serve as a background for the definitions of morality, moral behavior, and moral character which follow.

DEFINITIONS

Morality

According to Hart (1962), the basis of any morality is a system of rules of conduct which defines a network of reciprocal rights and obligations and which prohibits at least gross acts of malevolence. Moralities, then, are systems of rules that are external to people, designed to guide social or interpersonal behavior, and which may to some degree be codified and spelled out. This definition does not require a morality to be entirely compatible with or to further unequivocally the welfare of individuals subject to it, nor does it require that a particular morality be completely equitable when reviewed from a perspective outside the system in which it applies. In fact, the demands of justice or fair play in a given situation may occasionally run contrary to the prevailing moral code.

Moral Behavior

Moral behavior consists of actions carried out with regard to the rules that apply in a given social context (see also, Wright, 1971, p. 15). Obeying, disobeying, enforcing, subverting, learning, and using rules are all types of moral conduct or misconduct requiring explanation; misconduct is thus considered a subclass of moral behavior. For pragmatic and statistical reasons, one form of moral behavior (obedience or conformance) is usually salient. The most impressive single feature of everyday life is the degree to which people obey the norms of their culture. Even such self-consciously debauched men as Aliester Crowley, the Marquis de Sade, and Ceseare Borgia committed only a statistically insignificant number of deviant acts each day, and these were of a highly idiosyncratic and specialized nature, confined to limited subsystems of behavior. In the final analysis, moral behavior typically comes down to either following or disregarding a social rule of some sort; consequently a major problem for the psychology of moral conduct is to account for social compliance or noncompliance.

Moral Character

Freud pointed out that neurotic symptoms could be understood only by considering them within the context of the character structure of the person involved. Thus cigarette smoking in an anxiety neurotic might represent a means for gaining social approval, whereas it would be self-destructive behavior in a depressive. For Freud, then, neurotic character rather than neurotic symptoms became the primary subject matter. Similarly, within the context of moral psychology, delinquency, civil disobedience, and altruistic actions can be understood only by taking into account the character structure of the person involved. The concept of character closely resembles the notion of personality. Character, however, is usually defined in terms of those dispositions and traits that are subject to moral evaluation within a society. That is, character is defined not by what a person does, but by his reasons for doing it, by the recurring motives and dispositions that give stability and coherence to his social conduct. A second major problem for moral psychology, therefore, is to describe the dimensions along which character structure seems to vary.

THE MODEL

There are five recurring themes in philosophy and psychology which, when taken together, explain a considerable range of moral behavior and define certain important parameters of character development. These elements are at least conceptually independent and seem to characterize, in a formal way, how people differ in their use of rules. That is, these concepts are abstract dimensions of individual differences in nomotic (rule-governed) behavior, and as such, they should help to explain moral conduct in any sociocultural context. The first dimension is simply moral knowledge. The second corresponds to socialization in the Freudian sense of superego development, and is denoted by having an internal as opposed to an external viewpoint concerning the rules of one's society. The third dimension involves role taking and the recognition of the rights of others; it is the disposition to adopt the "moral point of view." The fourth dimension concerns the degree to which a person considers rules to be useful for regulating human affairs. The final dimension is defined by the degree of autonomy a person displays in his moral conduct.

As an aid to understanding the interrelationships among these five concepts, consider the following example as a paradigm case. A college student of modest talent, while taking a difficult examination whose outcome he believes will affect his future career, observes the answer sheet of an unusually gifted student lying in plain view in front of him. However, he deliberately ignores the answer sheet, thereby losing an opportunity to improve his score. How can we explain his action?

MORAL KNOWLEDGE

Explanatory Role of Moral Knowledge

The first answer to our paradigm question is that the student knows a rule that prohibits copying answers from another person's test. This is important because people can follow only the rules that they know. Moral knowledge means knowledge of moral rules, and a proper test of moral knowledge assesses the number and kinds of rules a person can state, or the variety of rules that he can correctly use. It does not determine how he feels about these rules, how he reasons with them, or how he evaluates them.

Because moral knowledge implies that moral learning has occurred, the next question concerns the kinds of things that are learned during the moral education of the child. Children seem to learn three distinct types of rules: (*a*) specific, primarily negative injunctions related to particular acts, such as "Don't get dirty," "don't hurt the cat;" (*b*) moral principles—general rules or norms of conduct, corresponding to Piaget's notion of a schema, that permit flexibility in social action and enable the child to choose between conflicting injunctions (e.g., the concept of reciprocity—cf. Gouldner, 1960); (*c*) comparison rules—cognitive strategies by which the person is able to compare what he is doing with an ideal norm of behavior. The first two rules are substantive while the third is procedural; that is, comparison rules (*c*) operate primarily on injunctions (*a*), while moral principles (*b*) mediate between them.

Moral knowledge is important because it provides the potential for self-control. That is, by learning a set of specific injunctions, moral principles, and comparisons rules, one acquires the cognitive foundation for self-criticism, which in turn is a necessary (but not sufficient) condition for the later development of self-control. Assume, for example, that a tennis player is losing a match and, after mentally comparing his strokes with the various rules of tennis form, he decides he is not volleying deeply enough. Whether or not he controls his subsequent strokes in accordance with this self-critical perception depends largely on his interest in winning. Thus, knowledge of the appropriate rules and the capacity for self-criticism afford the possibility but not the assurance of self-control.

Measurement of Moral Knowledge

Although considerable effort has been spent studying the behavioral implications of moral knowledge (cf. Pittel & Mendelsohn, 1966), this research has produced essentially negligible results. For example, the Hartshorne and May (1930) studies included several measures of moral knowledge, none of which predicted resistance to temptation in check situations.

Nor was there a significant relationship between a child's evaluation of the wrongness of cheating during interviews and his behavior in experimental situations. Similarly, Gordon, Short, Cartwright & Strodtbeck, (1963) found no differences in the degree to which six groups of teenage boys (black and white; delinquent and nondelinquent; lower class, middle class) understood the applicability of middle-class norms. Moreover, as Maller (1944) observed, tests of moral knowledge correlate about as highly with IQ tests as IQ tests correlate with themselves. Moral knowledge seems to be primarily related to intelligence, background, and desire to make a good impression (cf. Kohlberg, 1964). Because we are normally justified in assuming that people know the rules, it seems naive to assume a strong connection between knowledge of the rules and willingness to abide by them. As Malinowski (1932) pointed out in *Crime and Custom in Savage Society,* there are deviants in every society; however, no one ever doubts what is held to be right and wrong. The problem lies in learning to live with the "right" of one's society.

SOCIALIZATION

The Explanatory Role of Socialization

The second answer to the paradigm question is that our student may consider the rule about test cheating to be personally binding; that is, he has "internalized" the rules of test taking. The term socialization has few distinctive conceptual properties remaining after years of uncritical use (cf. Clausen, 1968). For the purposes of this paper, however, a person may be considered socialized to the degree that he regards the rules, values, and prohibitions of his society as personally mandatory. Conversely, to the extent that one feels estranged from the rules and norms of his social group, he tends to be unsocialized. Hart's (1962) distinction between an internal and an external attitude toward social rules reflects the phenomenology of socialization:

> When a social group has certain rules of conduct, this fact affords an opportunity for many closely related yet different kinds of assertions; for it is possible to be concerned with the rules, either merely as an observer who does not himself accept them, or as a member of the group which accepts and uses them as guides to conduct. We may call these, respectively, the "external" and the "internal" points of view [p. 86].

> Hence a society with law contains those who look upon its rulings from the internal point of view as accepted standards of behavior, and not merely as reliable predictions of what will befall them, at the hands of officials, if they disobey. But it also comprises those upon whom, either because they are malefactors or mere helpless victims of the system, these legal standards have to be imposed by force or threat of force; they are concerned with the rules merely as a source of possible punishment [p. 197].

Developmental Antecedents of Socialization

Most social psychologists assume some sort of natural antagonism between the individual and society. Freud, for example, believed that a child becomes able to live in society only after his natural but implacably anarchic tendencies are permanently repressed. One might equally hypothesize, however, that children are not anarchical but social by nature, that they enter the world programmed to be obedient, and that warmth and nurturance are essential in eliciting these tendencies. For example, in a sample of 11-month-old infants, Stayton, Hogan, and Ainsworth (1971) found that, of four key parent–child variables, only maternal warmth predicted obedience to commands ($r = .74$). Thus, rather than ask what must be done to the child to fit him into society, it may be more important to ask what must be done in order to drive him out.

Considerable research has been directed toward discovering the developmental antecedents of socialization, and the resulting literature is too extensive to review here. The available evidence, however, strongly suggests that while warmth is important, "love is not enough," that as children become more mobile, control factors play a correspondingly larger role in socialization. Thus warm, nurturant, *and* consistently restrictive or "authoritative" parents produce the most socialized children (cf. Bandura & Walters, 1959; Baumrind, 1971; Becker, 1964; Bronfenbrenner, 1970). The socialization process is largely completed by the time a child enters school, and results in what Piaget (1964) called "moral realism," that is, an authoritarian ethic of conformity.

Measurement of Socialization

It is with regard to the socialization dimension that a major measurement breakthrough in the study of moral conduct first occurred. The Socialization scale of the California Psychological Inventory (Gough, 1969; Gough & Peterson, 1952), an empirically keyed measure developed by comparing the responses of a large number of delinquents and nondelinquents, was specifically designed to assess the degree to which a person has internalized the rules, values, and conventions of his society. As an example of its remarkable properties, the measure was given in eight different languages in 10 countries to totals of 21,772 nondelinquents and 5,052 delinquents. In every comparison the test differentiated significantly between delinquents and nondelinquents; furthermore, in no instance was the mean of a delinquent sample in one country greater than the mean of a nondelinquent sample in any other country (cf. Gough, 1965). The scale seems to work equally well at various levels of socialization. At the upper level, Holland (1959) found, for example, that Socialization scores significantly predicted academic achievement among National Merit Scholars. At the lower end, Vincent (1961) found that the scale differentiated between female welfare recipients with one illegitimate child and those with two or more. Thus, Gough's Socialization scale seems to be a highly valid empirical index of the degree to which one has internalized culturally defined rules.

Willingness to follow rules is an important determinant of moral behavior and a major component of character development. Nonetheless, two considerations suggest that further information is necessary in order to assess adequately a person's character structure. First, other things being equal, well-socialized persons are also likely to be stuffy, rule-bound, pedantic prigs (e.g., "moral realists"). Second, many people who consistently comply with social norms did not enjoy the sort of childhood experiences that produce a high degree of socialization. As Megargee, Parker, and Levine (1971) demonstrate, within any population of nondelinquents, it is always possible to identify a sizable subgroup characterized by low Socialization scores. Phrasing the question in terms of our original paradigm case, how can we explain the student's refusal to cheat if we discover that he is in fact poorly socialized?

EMPATHY

Explanatory Role of Empathy

For at least 300 years philosophers in the Anglo-American (or utilitarian) tradition of ethics have assumed that man has an innate social sensitivity which plays an important role in moral development. John Stuart Mill suggested, for example, that man is naturally sympathetic and that a sense of duty depends on the development of these social feelings. Kurt Baier, a modern utilitarian philosopher, has proposed that certain social rules are justified when seen from the "moral point of view," a perspective which tends to promote the common good. This viewpoint is a natural attitude which may be taught to children: "When we teach children the moral point of view, we try to explain it to them by getting them to put themselves in another person's place: How would you like to have that done to you [Baier, 1965, p. 107]!" Thus, when a person acts from the moral viewpoint, he tries to consider the implications of his actions for the welfare of others. The disposition to take the moral point of view is closely related to empathy or role taking.

Several social psychologists, including J. M. Baldwin, W. McDougall, and G. H. Mead, have also assumed that people have an innate capacity for empathic sensitivity which is normally elicited by social experience. For these role theorists, group participation and an increased sensitivity to the demands of others results in self-consciousness which, in turn, causes the child to observe the rules of his society. The implications of all this for the explanation of social compliance is that the test-taking student in our paradigm case may refrain from cheating because he is sensitive to his peers' expectations concerning test behavior; cheating might equally be seen as taking unfair advantage of the other students.

Measurement of Empathy

Hogan (1969) developed a 64-item, empirically keyed empathy scale by comparing the responses of 211 subjects, divided into high- and low-subgroups on the basis of empathy ratings, across 957 items taken from the California Psychological Inventory, the MMPI, and a special pool of additional items. The best item on the scale in the original item analysis was: "As a rule I have little difficulty in 'putting myself into other people's shoes.'" The scale correlated .00 with Socialization, indicating that the two dimensions are indeed independent. It also correlated .58 with rated social acuity, .53 with overall skill at playing charades, and .43 with rated ethical sensitivity.

If socialization and empathy are independent aspects of moral development, then individual differences along these two dimensions should be associated with differences in character structure and moral conduct (see Table 1). Specifically, other things being equal, persons scoring low on both scales tend to be delinquent. Those receiving low scores for empathy but high scores for socialization tend to be rigid rule followers—Piaget's moral realists. Persons with low scores for socialization but high scores for empathy tend to be cavalier about the conventional rules of society; they are "emancipated," mildly sociopathic members of normal society, that is, persons who double-park in parking lots, do not return borrowed books, and smoke marijuana —Piaget refers to them as "chic types." Persons with high scores for socialization and empathy tend to be morally mature; their compliance with social rules is effortless but their attitudes are tempered by a sympathy for the moral frailities of others.

It is empirically as well as conceptually useful to assess social conduct in terms of both socialization and empathy. For example, the mean scores for socialization and empathy for a sample of 100 inmates of a New York State reformatory were 26.2 and 29.1.[4] Comparable values for a group of 100 Air Force officers were 36.4 and 37.7. By considering all persons scoring below 36.5 on socialization and 39.0 on empathy as delinquent, 95% of the delinquents were accurately classified, while 27% of the nondelinquents were misclassified. In a sample of 594 male undergraduates, Kurtines and Hogan (1972) identified 130 students whose socialization scores were less than or equal to the mean socialization score of 119 incarcerated delinquents ($X = 26.88$). As expected, the unsocialized students were relatively empathic compared to the delinquents ($r_{\text{bis}} = .44$). In a study of undergraduate marijuana use (Hogan, Mankin, Conway, & Fox, 1970), students were placed in three groups: users, nonusers, and principled nonusers(e.g., students who said that they had not smoked marijuana and never would). Relative to men in general, users received low scores for socialization and high scores for empathy, while the pattern was precisely reversed for the principled nonusers. Thus there is some evidence for the characterological implications suggested by Table 1.

TABLE 1

SUGGESTED CHARACTEROLOGICAL IMPLICATIONS OF THE INTERACTION BETWEEN SOCIALIZATION AND EMPATHY

Empathy	Socialization	
	Low	High
High	"Le chic type"	Morally mature
Low	Delinquent	"Moral realist"

Developmental Antecedents of Empathy

In spite of the importance of empathy or role-taking ability as an explanation for moral conduct, little is known about the antecedents of this disposition. However, it seems reasonable to assume that the following four factors are related to the development of empathy. First, Mead (1934) thought role-taking ability was the "g" factor in intelligence. Role taking requires that a person adopt alternative perspectives with regard to various issues; the process of shifting perspectives seems similar to spatial reasoning (an important component of intelligence); and the empathy scale correlates between .30 and .50 with several measures of intellectual performance. Thus there seems to be some association between intelligence and empathy. Second, being required to adopt alternative perspectives *vis-á-vis* one's parents should facilitate the development of role-taking skills. Consequently, parents who either overindulge or consistently reject their children probably fail to stimulate their children's natural empathic tendencies. Third, intelligence and practice at role taking are necessary but insufficient conditions to produce empathic behavior—the child must also be willing to act on his empathic perceptions.

---

[4] These data were supplied by Professor H. G. Gough and the Institute of Personality Assessment and Research at U. C. Berkeley, whose cooperation I gratefully acknowledge.

Such a willingness probably depends on two kinds of experience: (*a*) receiving empathic treatment at some time in one's life and (*b*) enduring sufficient injustice, ridicule, betrayal, or persecution that the experience makes a lasting impression. The fourth factor that seems to contribute to an empathic disposition is a relative absence of repression or denial—an openness to inner experience, a willingness to attend to intuitive promptings and nonverbal cues. The development of empathy, like cognitive development generally, is probably finished by late adolescence and results in what may be called an "ethic of reciprocity," characterized by cooperation and peer group conformity.

In spite of the obvious importance of an empathic disposition in the formation of moral character, it is neither the only nor the most important factor in the process. There are two reasons for this. First, interviews with subjects who receive very high scores on the empathy scale suggest that they often suffer from an excess of role taking: They are too concerned with the expectations of others, they excessively inhibit hostility and aggression, and they suffer from identity diffusion. Unleavened role taking can produce an equivocating jellyfish as well as a compassionate person with a broad moral perspective. Furthermore, many persons whose behavior accords with social norms and conventions have experienced none of the developmental circumstances necessary to elicit an empathic disposition. To return to our paradigm case, how might we account for the student's reluctance to cheat if we discover that he is both unsocialized and nonempathic?

ETHICS OF CONSCIENCE AND THE ETHICS OF RESPONSIBILITY

Explanatory Role of the Ethics of Conscience and the Ethics of Responsibility

There is a branch of moral philosophy which analyzes and evaluates arguments used to justify social and legal institutions. It considers questions of the following type: "To what kinds of considerations is appeal properly made when the authority of the political and legal order and our obligations to comply with its dictates are called into question [Olafson, 1961]?" In the history of modern legal philosophy two forms of justification continually recur. Advocates of the first point of view argue that there are higher laws, unrelated to human legislation, which may be discovered by intuition and reason; a human law is just if, and only if, it corresponds to or can be derived from the higher laws. Those who take the second viewpoint deny the existence of "higher" laws. Instead, they justify their arguments in terms of the instrumental value of the manifest law as a means for promoting the general welfare of society. Laws and political institutions are merely instruments for the realization of the common good; just laws are those which tend on the whole to maximize happiness.

With certain exceptions, the first position has traditionally had the greater influence:

> There can be little doubt that the system of justification which has exerted the widest intellectual influence and most decisively shaped the further development of social philosophy is the theory of natural [i.e., higher] law. It may even be said that with the idea of natural law the whole enterprise of calling existing institutions to moral accounting gets underway [Olafson, 1961, p. 8].

Utilitarianism and legal positivism emerged in the eighteenth century as a reaction against the theory of higher law. The utilitarian philosophers, primary architects of English social reform in the nineteenth century, thought that the concept of higher law was an arbitrary standard for appraising institutions, and that it encouraged an anarchical individualism inconsistent with the rule of law and settled society.

The intellectual and psychological postures underlying the higher law and positive law traditions in moral philosophy have numerous counterparts in everyday life and represent a relatively unexplored dimension of character development. The everyday equivalent of higher law morality can be called the ethics of personal conscience, while the ethics of social responsibility reflects the positive law tradition. The ethics of conscience is the dominant viewpoint of the contemporary American intellectual establishment; it is represented, for example, by many contributors to the *New York Review of Books*. On the other hand, Supreme Court Justice Frankfurter, Erwin Griswold, former dean of Harvard Law School and Solicitor General of the United States, and H. L. A. Hart, the distinguished English legal and moral philosopher represent prominent recent spokesmen for the ethics of responsibility.

Consider once again our hypothetical, noncheating student whom we have discovered is neither well-socialized nor empathic. If he advocates the ethics of responsibility, he may then comply with the rules because they make his world more orderly and predictable. That is, a belief in the instrumental value of rules should produce social compliance despite deficient socialization and insensitivity to the expectations of others. Conversely, persons who are nonempathic, unsocialized, and endorse the ethics of conscience will very likely be delinquent.

Measurement of the Ethics of Conscience and the Ethics of Responsibility

Hogan (1970) developed a measure of the disposition to adopt one or the other of these viewpoints. There are two forms of the test, called the Survey of Ethical Attitudes, each with 35 items. The following statements are representative: (*a*) All civil laws should be judged against a higher moral law; (*b*) right and wrong can be meaningfully defined only by the law; (*c*) an unjust law should not be obeyed; (*d*) without law the life of man would be nasty, brutish, and short.

These items are answered by checking one of five response options ranging from agree strongly to disagree strongly. The parallel form reliability of the test in a sample of 98 college men was .88. The measure is uncorrelated with intelligence, and in two separate samples it discriminated very strongly between persons whose vocational choice reflected a belief in law and established procedures (i.e., policemen, ROTC seniors) and persons who believed in civil disobedience as a means for promoting social change (cf. Hogan, 1970).

In contrast with Kohlberg (1963) and Piaget (1964), the author assumes that moral conduct is fundamentally "irrational," that differences in even such obviously cognitive phenomena as moral judgments derive from more basic personological structures. If so, then there should be clear-cut personality correlates of the disposition to adopt the ethics of conscience or responsibility. Correlations with the California Psychological Inventory, the Myers-Briggs Type Indicator (Myers, 1962), and several other personality measures suggest that persons who adopt the ethics of personal conscience possess an interesting set of positive and negative characteristics. They are independent, innovative, and form creating; however, they also tend to be impulsive, opportunistic, and irresponsible. From peer descriptions the adjective with the highest correlation (.49) with the ethics of conscience was "rebellious." Persons whose moral judgments reflect the ethics of social responsibility also show a blend of attractive and disagreeable features. On the one hand, they are reasonable, helpful, and dependable; on the other hand, they often seem conventional and resistant to change. The single adjective from peer descriptions most characteristic ($r =$ .47) of the ethics of responsibility was "thoughtful."

Developmental Antecedents of the Ethics of Conscience and the Ethics of Responsibility

The next question concerns the developmental antecedents of the ethics of conscience and responsibility. A study by Hogan and Dickstein (1972b) provides some insight into this problem. Interviews with high and low scorers on the Survey of Ethical Attitudes suggested that both groups were highly sensitive to injustice, but differed markedly with regard to their perceptions of the sources of injustice in society. The ethics of conscience seemed related to the belief that people are naturally benevolent and that social injustice is produced by dehumanizing and oppressive institutions; the ethics of responsibility seemed to rest on the notion that people are naturally malevolent and that institutions restrain the implacably antisocial impulses of men. Scores for the tendency to blame people or institutions were assigned on the basis of a projective test (Hogan & Dickstein, 1972a). The average correlation between these scores and the Survey of Ethical Attitudes across three groups (total $N = 92$) was .58 (.77 when corrected for attenuation), suggesting that the ethics of conscience and responsibility are indeed related to a person's perceived locus of injustice. Thus home environments which encourage a belief in natural human goodness and a suspicion of institutions tend to promote the ethics of conscience. Conversely, parents who foster skepticism concerning the motives of others encourage a belief in the instrumental value of law and, consequently the ethics of responsibility (see also MacDonald, 1971). Finally, while the moral and developmental implications of the ethics of conscience and responsibility can be properly understood only within the context of a person's total character structure, the most mature people should tend to cluster in the center of this dimension.

AUTONOMY

Explanatory Role of Autonomy

Two considerations suggest that the dimensions of moral development discussed this far (i.e., moral knowledge, socialization, empathy, and the ethics-of-conscience—ethics-of-responsibility continuum) are, by themselves, incomplete either as explanations of moral conduct or descriptions of character structure. First, from a philosophical perspective, these dimensions reflect a distinct utilitarian bias. However, Kant argued that compliance with social norms based on self-interest (or the larger welfare of society) is in no sense moral. Rather, the truly moral man has an autonomous will and governs his actions by a personal sense of duty. Reflecting the Kantian ethic, Nietzsche described utilitarianism as a morality fit only for English shopkeepers. Thus, from the viewpoint of traditional ethics, a complete description of moral development must include the factor of autonomy. The discussion of character development presented thus far also appears inadequate when seen from the perspective of common sense. That is, a well-socialized and empathic adherent of either the ethics of conscience or responsibility is socially conforming, and is a model citizen when judged in terms of the standards of his sociocultural milieu. However, social and legal institutions are never totally just, and political leadership is often misguided (if not actually corrupt). Consequently, it sometimes happens that a model citizen may act immorally as a result of complying with collective norms. The development of an autonomous set of moral standards serves to insulate one from the potential immorality of the community.

Returning for the final time to our noncheating student, he may refuse to copy not because he is socialized, not because he is concerned about the evaluations of his peers, not because he feels rules are useful, but because he considers cheating to be beneath his dignity as a person.

Autonomy is a troublesome concept. An autonomous person who is also unsocialized, nonempathic, and disregards the pragmatic value of rules as well, is likely to be autocratic and anticonforming, a great rogue and scoundrel. Obviously this dimension, like all the others, can never be considered in isolation.

Measurement of Autonomy

As a quantitative index of autonomy, the author uses a measure of independence of judgment developed by Barron (1953). This 22-item scale was constructed by comparing the responses of yielders and nonyielders in Asch's original (1956) conformity studies. The measure has provided some positive results; however, the items are quite complex and the resulting scores have only marginal reliability. An alternative measure of autonomy with simpler items and improved reliability would be a useful measurement contribution, and work is currently proceeding on such a measure.

Although social psychologists have spent a great deal of time studying suggestibility and conformity, not much is known about autonomy. In an unpublished study, Kurtines[5] gave the California Psychological Inventory to 30 fraternity members who were undergraduates at Johns Hopkins University. Eleven of these persons rated all the others for autonomy using a 7-point scale. Each person was assigned an autonomy score based on the average of the 11 ratings he received; the average interrater correlation was .41, and the reliability of the composite autonomy ratings was .89. Eight of the 18 California Psychological Inventory scales correlated above .39 ($p < .05$) with these ratings; they were: Dominance (.56); Capacity for Status (.60); Social Presence (.42); Self-Acceptance (.52); Good Impression (.47); Communality (.39); Achievement via Conformance (.40); and Intellectual Efficiency (.57). To estimate the pattern of personological variables most related to autonomy, a stepwise regression analysis was conducted, selecting five variables. The resulting equation, which had a multiple r of .76, included Dominance, Capacity for Status, and Self-Control with positive weights, Sociability and Femininity with negative weights. Although the size of this sample permits only tentative conclusions, the pattern of zero-order correlations with the California Psychological Inventory and the regression equation suggests that the autonomous person is strong, forceful, and self-ascendant; he manages his affairs very carefully, and is little affected by others in choosing or achieving his goals. Such a person resembles an Old Testament patriarch or Melville's Captain Ahab.

Developmental Antecedents of Autonomy

In *The Moral Judgment of the Child,* Piaget (1964) suggested that children's morality is initially "heteronomous," authoritarian, and conditioned upon a unilateral respect for adult authority. If left alone by adults, the child acquires through the natural processes of cognitive development and peer interaction an "autonomous" and democratic morality based on mutual respect among peers. On this point Piaget seems surely to have been wrong. Comparing child-rearing practices in the Soviet Union and the United States, Brofenbrenner (1970) observes, for example, that exclusive peer interaction in American children has produced effects precisely the opposite of those Piaget predicted; while American children, relative to Russian children, become emancipated from adult authority earlier, they remain thereafter excessively dependent on peer approval (see also Boehm, 1957).

In a carefully documented monograph, Baumrind (1971) suggests some important developmental precursors of autonomy. First, parents who are permissive, undemanding, passively accepting, or overprotective inhibit the development of autonomy in their children. Second, early development of cognitive and linguistic skills and stimulation of interest in school achievement seem to enhance a child's competence, and therefore, the self-esteem necessary for autonomous behavior. Third, parents who are themselves individualistic, independent, and demanding provide their children with the strongest models for autonomous conduct. Finally, parents who clearly label certain actions as praiseworthy or blameworthy, explain their rules and encourage verbal give and take, also promote the child's ability to "order" and control his own behavior. Although Baumrind has clarified the developmental antecedents of autonomy, this disposition probably does not fully emerge until a person leaves his peer group and begins his life (defined in terms of marriage and career). Then, in combination with socialization and empathy, autonomy may produce an "ethics of equity"—in psychological terms, moral maturity.

COMPLETE MODEL

Five important aspects of character development have been described which, it has been suggested, are related to five considerations required for the explanation of moral behavior. Each dimension is defined in terms of a specific assessment device. While the dimensions are conceptually independent, the relationships among the scales used to define them can be determined empirically. Table 2 is based on data taken from a variety of sources, and represents the best current

---

[5] W. Kurtines. A study of autonomy. Department of Psychology, Johns Hopkins University, October, 1972.

TABLE 2

ESTIMATED RELATIONSHIPS AMONG FIVE DIMENSIONS OF CHARACTER STRUCTURE

Dimensions	Correlations			
	So	Em	SEA	Autonomy
Moral knowledge	.00	.30	.00	.30
Socialization (So)	—	.00	.26	.09
Empathy (Em)		—	−.20	.41
Ethics of responsibility (SEA)			—	−.33

Note. SEA = Survey of Ethical Attitudes.

estimate of the magnitude of these relationships. The correlations with moral knowledge for socialization, empathy, and the Survey of Ethical Attitudes are based on associations with the Scholastic Aptitude Test. The correlation between autonomy and moral knowledge is an estimate derived from Crutchfield's (Krech, Crutchfield, & Ballachey, 1962, p. 526) report of a dependable relationship between IQ and nonyielding in his experiments. The other correlations summarize data from the author's files. No claims are made for the generality of these coefficients, and other investigators may want to determine their size for themselves.

These dimensions also work moderately well together. Hogan and Dickstein (1972a) gave three fraternities ($N = 92$) a 15-item sentence completion test on which subjects were asked to respond quickly and briefly to such statements as: (*a*) I think it is unnecessarily cruel to keep condemned prisoners on death row for so long, and to make the execution such an elaborate ritual. (*b*) The police should be encouraged in their efforts to apprehend and prosecute homosexuals. Homosexuality threatens the foundations of our society.

Responses to these items were scored for overall maturity of moral judgment using the following scoring elements: (*a*) concern for the sanctity of the individual; (*b*) judgments based on the spirit rather than the letter of the law; (*c*) concern for the welfare of society as a whole; (*d*) capacity to see both sides of an issue.

Four raters scored each protocol; 2 points were assigned to an item if any of the four scoring elements was clearly present in the response, 1 point if any of the scoring elements could be readily inferred, 0 if they were missing. The interrater reliability was estimated to be .88, and the reliability of the test itself was .82. The validity of this projective measure was estimated as follows: Each member of the fraternity rated the others for "sensitivity to injustice," defined as

> . . . one who is quick to perceive unfairness in the decisions of persons or groups, or in the treatment that persons or groups receive from others. He may or may not openly express his concern; however his feelings will be obvious to those who know him. One who is insensitive to injustice will be less likely to notice the unfairness in a situation, and will rarely show concern when such unfair treatment is pointed out to him.

The reliability of these ratings after refinement by factor analysis varied between .86 and .93. The correlations between scores for maturity of moral judgment and rated sensitivity to injustice were .43, .13, and .52 for the three fraternities, respectively.

If scores for maturity of moral judgment are taken as an index of "moral maturity," then correlations between these scores and the five dimensions of character development provide some indication of how well the total model works. These data are presented in Table 3, which contains no information about the relationship between mature moral judgment and moral knowledge. Kohlberg (1964) reports, however, that for a population of 72 adolescent boys the correlation between IQ and maturity of moral judgment (with age controlled) was .31. With the exception of the Survey of Ethical Attitudes, the correlations in Table 3 support the model well; for example, persons who make mature moral

TABLE 3

CORRELATIONS BETWEEN FOUR DIMENSIONS OF CHARACTURE STRUCTURE AND MATURITY OF MORAL JUDGMENT

Dimension	Mature moral judgment
Socialization	.40
Empathy	.58
Ethics of responsibility (SEA)	−.40
Autonomy	.56

Note. $N = 41$. SEA = Survey of Ethical Attitudes.

judgments tend to be well socialized, empathic, and autonomous. The negative correlation with the ethics of responsibility, however, requires interpretation. High scores on the Survey of Ethical Attitudes reflect the ethics of responsibility; low scores indicate the ethics of conscience. The mean score on the Survey of Ethical Attitudes for this sample was 14.3, and the range was 4–25. Ratings for mature moral judgment varied between 0 and 14.0. An examination of the bivariate plot reveals that the two persons with the highest (ethics of responsibility) and the two with the lowest (ethics of conscience) scores on the survey together received an average moral maturity rating of 4.0. On the other hand, the person with the highest moral maturity rating in the sample received a score of 18 on the survey. Thus, the scatter plot generally supports the prediction derived from the model that persons in the middle of the ethics-of-conscience–ethics-of-responsibility dimension are more mature than the persons at either extreme.

CHARACTEROLOGICAL IMPLICATIONS

Specifying dimensions of character development leads inexorably to a discussion of character types. Psychologists have proposed a number of character typologies which, while varying considerably in detail, also have many features in common (cf. Kay, 1968; Wright, 1971). In *The Psychology of Character Development*, Peck and Havighurst (1960) presented what is perhaps the most important of these; it might be useful to trace the parallels between their typology and the model of character development presented here. To define empirically the primary dimensions of character development, Peck and Havighurst (1960) obtained in situ ratings of adolescent subjects on 35 "moral" traits. These traits were intercorrelated, factor analyzed, and the three factors which emerged should be familiar. The first describes

> . . . behavior which conforms positively to the established moral code . . . a person who exemplified this vector is one who is conventionally moral, socially conforming to both the maternal and peer codes of approved behavior, and shows no hostility or emotional discomfort in doing so. . . . We might infer that the person high on this characteristic conforms, not so much from pressures to do so, but because he enjoys it and finds it the most comfortable and satisfying way to live [pp. 244–245].

This dimension is clearly socialization.

Peck and Havighurst's (1960) second dimension was defined by (*a*) the "degree to which subject accurately recognizes the usual behavior patterns of the people around him, can predict what they are likely to do, and how they are likely to react in a given action [p. 221]"; (*b*) the "capacity to understand other people's reasons for acting as they do [p. 221];" (*c*) the "capacity for 'feeling with' other people; for experiencing, at least to some extent, the same emotions they are experiencing at the moment [p. 221];" and (*d*) the "degree to which conscious or preconscious self-perception is congruent with the behavioral self [p. 223]." This second dimension closely approximates empathy.

The third dimension was more ambiguous. It was defined as ". . . the degree to which behavior is directed by, or is in accord with a present and functioning superego [p. 236]." This depends on

> . . . the degree to which superego directives are integrated into behavior, i.e., a person with only self-punitive superego "voices" which do not forestall immoral behavior does not have an integrated system of principles . . . guilt per se is not a direct measure of superego strength . . . [Peck & Havighurst, 1960, p. 236].

The third dimension resembles what we have called autonomy.

Peck and Havighurst next define five character types, labeled amoral, expedient, conforming, irrational-conscientious, and rational-altruistic. The first represents a person who lacks any self-imposed control or concern for the moral requirements of social living. However, the ". . . last four represent the four kinds of reasons why a person may behave according to the moral standards of his society [p. 4]." And the development of these character types is explained in terms of the dimensions we have called socialization, empathy, and autonomy (cf. Peck & Havighurst, 1960, p. 87). That is, expedient types were intelligent enough to understand the utility of conformance; however, they were unsocialized, nonempathic, and nonautonomous. Conforming subjects were well socialized but neither empathic nor autonomous. Those classified as irrational-conscientious were well socialized, autonomous, and nonempathic. Only the few rational-altruistic persons were socialized, empathic, and autonomous.

Let us now review the complete model by considering in what sense it can be called developmental. Two themes define the developmental perspective. The first is that there are definable endpoints to development, endpoints which are preceded by qualitative changes over time. According to the model presented here, the *ideal* endpoint of moral development is moral maturity, defined in terms of optimal placement on the five dimensions; the *practical* endpoint is social conformance, which can be explained in terms of these dimensions. (The relationship between the two endpoints is seen in the fact that while the morally mature typically conform to rules of conduct, all those who conform are not necessarily mature.)

Moreover, the emergence of socialization, empathy, and autonomy represent transition points which are followed by qualitative changes in the dynamics of social behavior. That is, until a child becomes socialized, he is egocentric, impulsive, and undisciplined. After socialization but before empathy develops, he is excessively respectful of adult authority. An empathic but nonautonomous person places a greater priority on human needs than on the maintenance of rules; however his conduct remains closely tied to the expectations of his peer group. It is only after a degree of autonomy has been achieved that behavior may become independent of external controls. In contrast with these three dimensions, moral knowledge and the ethics-of-conscience–ethics-of-responsibility continuum are not developmental in the sense described here.

The second theme defining the developmental perspective is that life is a continuous process of adjusting internal conditions to external demands; these adjustments are made necessary by changes which occur in both the organism and the environment. The present model also incorporates this characteristic of developmental theory. The major problem for an infant is securing adequate nurturance and protection from his caretakers. Moreover, as the child becomes mobile, he tends to wander further from his mother's protection, thereby exposing himself increasingly to danger. Thus obedience to maternal commands both facilitates mother–child interaction and improves the child's chances for survival. The primary developmental task for a child in the middle years is to become a social being, a process brought about almost exclusively through peer interaction. For the typical child such interaction begins around 4 or 5 years of age, where because adaptation to the group requires accommodation to a radically expanded set of norms, continued strict obedience to parental commands is maladaptive. The development of empathy greatly improves the capacity to anticipate and the willingness to comply with the expectations of the group—the prerequisite for sociability—and increases flexibility in peer interaction. Finally, the most important problem for the young adult is to establish his own life. Yet most young people are caught between the internalized demands of their family and the external expectations of the peer community. Thus the development of autonomy is an important solution to the functional dialectic of development; it allows one to deviate from both parents and peers, but in a (hopefully) prosocial fashion. Seen in this way, the emergence of socialization, empathy, and autonomy are adaptive within the context of the social environment during maturation; they are capacities which mediate both the needs of the individual and the demands of his social group.

The present model differs from conventional developmental theories in one important respect. Most such theories describe development as proceeding through preset stages in a step-wise fashion, and successive levels can be achieved only after the preceding stages have been mastered. This paper suggests that socialization, empathy, and autonomy are major transition points in moral development which occur at progressively later points in time. Moreover, once attained, these capacities bring about qualitative changes in the underlying structure of moral conduct (which phenotypically may remain unchanged). In this model, however, attainment of the later "stages" is *not* dependent on successful transition through the earlier levels. Rather, all three stages are distinct developmental challenges whose outcome defines each person's unique character structure.

Because contemporary social theorists often seem unconcerned with the problem of conformity, one last comment on this topic is appropriate. The model presented here defines the formal parameters of moral maturity. However, these qualities may rarely exist in a single person. If so, then the obvious conclusion is that most people are less than morally mature. Furthermore, the proportion of thugs and psychopaths in society seems independent of their place in the social order; social class determines the form rather than the frequency of delinquency conduct. In addition, evidence from history suggests that to expect people to be compassionate, charitable, and tolerant is patently unrealistic. As Camus (1955) observed, the "Be virtuous" of our confessionals reveals "a nostalgia at the same time as an ignorance;" it is a sterile exercise on a great subject. Paralleling Camus, Freud persistently maintained in his later writings that the most we can reasonably expect is that people will comply with the rules, that they will avoid doing evil, that they will not steal our property, molest our families, or enslave us. Consequently, since little beyond that seems possible, the practical endpoint of moral development, social conformance, must

REFERENCES

ASCH, S. Studies of independence and conformity. *Psychological Monographs,* 1956, **70**(9, Whole No. 416).

BAIER, K. *The moral point of view.* New York: Random House, 1965.

BANDURA, A., & WALTERS, R. H. *Adolescent aggression.* New York: Ronald Press, 1959.

BARRON, F. Some personality correlates of independence of judgment. *Journal of Personality,* 1953, **21**, 789–797.

BAUMRIND, D. Current patterns of parental authority. *Developmental Psychology,* 1971, **4**(1, Pt. 2).

BECKER, W. C. Consequences of different kinds of parental discipline. In M. Hoffman & L. W. Hoffman (Eds.), *Review of child development research.* Vol. 1. New York: Sage Foundation, 1964.

BOEHM, L. The development of independence: A comparative study. *Child Development,* 1957, **28**, 85–92.

BRONFENBRENNER, U. *Two worlds of childhood.* New York: Sage Foundation, 1970.

CAMPBELL, D. G. Ethnocentric and other altruistic motives. In D. Levine (Ed.), *Nebraska Symposium on motivation: 1965.* Lincoln: University of Nebraska Press, 1965.

CAMUS, A. *The myth of sisyphus.* New York: Anchor, 1955.

CLAUSEN, J. A. (Ed.) *Socialization and society.* Boston: Little, Brown, 1968.

DARWIN, C. *The expression of emotions in man and animals.* London: John Murray, 1872.

GOFFMAN, E. *The presentation of self in everyday life.* Garden City, N. Y.: Doubleday, 1959.

GORDON, R., SHORT, J., CARTWRIGHT, D., & STRODTBECK, F. Values and gang delinquency. *American Journal of Sociology,* 1963, **69**, 109–128.

GOUGH, H. G. Conceptual analysis of psychological test scores and other diagnostic variables. *Journal of Abnormal Psychology,* 1965, **70**, 294–302.

GOUGH, H. G. *Manual for the California Psychological Inventory.* (Rev. ed.) Palo Alto, Calif.: Consulting Psychologists Press, 1969.

GOUGH, H. G., & PETERSON, D. R. The identification and measurement of predispositional factors in crime and delinquency. *Journal of Consulting Psychology,* 1952, **16**, 207–212.

GOULDNER, A. W. The norm of reciprocity: A preliminary statement. *American Sociological Review,* 1960, **25**, 161–179.

HAMILTON, W. D. Selection of selfish and altruistic behavior in some extreme models. In J. F. Eisenberg & W. S. Dillon (Eds.), *Man and beast: Comparative social behavior.* Washington, D. C.: Smithsonian Institute Press, 1971.

HART, H. L. A. The concept of law. Oxford: Oxford University Press, 1962.

HARTSHORNE, H., & MAY, M. A summary of the work of the Character Education Inquiry. *Religious Education,* 1930, **25**, 607–619.

HOGAN, R. Development of an empathy scale. *Journal of Consulting and Clinical Psychology,* 1969, **33**, 307–316.

HOGAN, R. A dimension of moral judgment. *Journal of Consulting and Clinical Psychology,* 1970, **35**, 205–212.

HOGAN, R., & DICKSTEIN, E. A measure of moral values. *Journal of Consulting and Clinical Psychology,* 1972, **39**, 210–214. (a)

HOGAN, R., & DICKSTEIN, E. Moral judgment and perceptions of injustice. *Journal of Personality and Social Psychology,* 1972, **23**, 409–413. (b)

HOGAN, R., & HENLEY, N. Nomotics: The science of human rule systems. *Law and Society Review,* 1970, **5**, 135–146.

HOGAN, R., MANKIN, D., CONWAY, J., & FOX, S. Personality correlates of undergraduate marijuana use. *Journal of Consulting and Clinical Psychology,* 1970, **35**, 58–63.

HOLLAND, J. L. The prediction of college grades from the California Psychological Inventory and the Scholastic Aptitude Test. *Journal of Educational Psychology,* 1959, **50**, 135–142.

HUIZINGA, J. *Homo ludens.* Boston: Beacon Press, 1955.

Kay, W. Moral development. London: Allen & Unwin, 1968.

Kohlberg, L. The development of children's orientations toward a moral order. *Vita Humana,* 1963, 6, 11–33.

Kohlberg, L. Development of moral character and ideology. In M. L. Hoffman & L. W. Hoffman (Eds.), *Review of child development research.* Vol. 1. New York: Sage Foundation, 1964.

Krech, D., Crutchfield, R. S., & Ballachey, E. L. *Individual in society.* New York: McGraw-Hill, 1962.

Kurtines, W., & Hogan, R. Sources of conformity in unsocialized college students. *Journal of Abnormal Psychology,* 1972, **80**, 49–51.

Lorenz, K. *On aggression.* New York: Harcourt Brace, 1966.

MacDonald, A. P. Jr. Correlates of the ethics of personal conscience and the ethics of social responsibility. *Journal of Consulting and Clinical Psychology,* 1971, **37**, 443.

Maller, J. B. Personality tests. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald Press, 1944.

Malinowski, B. *Crime and custom in savage society.* New York: Harcourt, Brace, 1932.

McDougall, W. *Social psychology.* London: Methuen, 1908.

Mead, G. H. *Mind, self, and society.* Chicago: University of Chicago Press, 1934.

Megargee, E. I., Parker, G. V. C., & Levine, R. V. Relationship of familial and social factors to socialization in middle-class college students. *Journal of Abnormal Psychology,* 1971, **77**, 76–89.

Myers, I. B. *Manual* (1962), *the Myers-Briggs Type Indicator.* Princeton: Educational Testing Service, 1962.

Olafson, F. A. *Society, law, and morality.* Englewood Cliffs: Prentice-Hall, 1961.

Peck, R. F., & Havighurst, R. J. *The psychology of character development.* New York: Wiley, 1960.

Peters, R. S. *The concept of motivation.* London: Routledge & Kegan Paul, 1958.

Piaget, J. *The moral judgment of the child.* New York: Free Press, 1964.

Pittel, S. M., & Mendelsohn, G. A. Measurement of moral values. *Psychological Bulletin,* 1966, **66**, 22–35.

Stayton, D., Hogan, R., & Ainsworth, M. D. S. Infant obedience and maternal behavior: The origins of socialization reconsidered. *Child Development,* 1971, **42**, 1057–1069.

Tinbergen, N. On war and peace in animals and man. *Science,* 1968, **160**, 1411–1418.

Vincent, C. *Unmarried mothers.* New York: The Free Press, 1961.

Waddington, C. H. The ethical animal. Chicago: University of Chicago Press, 1967.

Wright, D. *The psychology of moral behavior.* Baltimore: Penguin Books, 1971.

(Received May 23, 1972)

Manuscripts Accepted for Publication in the *Psychological Bulletin*

On the Subjective Character of the Empirical Base of Guilford's Structure-of-Intellect Model. J. L. Horn (Department of Psychology, University of Denver, Denver, Colorado 80210).

Tactile Communication of Speech: A Review and an Analysis. Jacob H. Kirman (Department of Psychology, Queens College, Flushing, New York 11367).

Cross-Cultural Research in the Perception of Pictorial Materials. Robert J. Miller (Department of Psychology, Pennsylvania State University, University Park, Pennsylvania 16802).

Comment on Regression Analysis of Proportional Cell Data. John E. Overall (Department of Neurology and Psychiatry, University of Texas Medical Branch, Galveston, Texas 77550).

What the Mind's Eye Tells the Mind's Brain: A Critique of Mental Imagery. Zenon W. Pylyshyn (Department of Psychology, University of Western Ontario, London, Ontario, Canada 72).

Can a Personality Construct be Validated Empirically? Donald W. Fiske (Department of Psychology, University of Chicago, Chicago, Illinois 60637).

Psychological Bulletin
1973, Vol. 79, No. 4, 233–238

ELECTRODERMAL ACTIVITY AS AN INDEX OF AROUSAL IN SCHIZOPHRENICS[1]

RICHARD A. DEPUE[2]
Southern Illinois University at Carbondale

DON C. FOWLES
University of Iowa

Previous reviews of research utilizing electrodermal activity as an index of arousal in schizophrenia have found no consistent differences between schizophrenics and normals, although other psychophysiological measures have shown chronic schizophrenics to be more aroused. A more detailed look at the electrodermal research, however, reveals that consistent results have been reported for two aspects of electrodermal activity: (*a*) habituation of responses or levels during periods of minimal stimulation, such as tones and lights of low intensity and (*b*) frequency of spontaneous electrodermal responses regardless of stimulus conditions. Both of these measures appear to reflect arousal and yield results indicating that chronic schizophrenics are overaroused.

Since the earlier writings of Duffy (1951), Lindsley (1951), Hebb (1955), and Malmo (1959), the construct of arousal (or activation) has gained enormous impetus as an explanatory concept in theories of behavior and has provided a framework for research and theory in the area of psychophysiology (Lacey, 1967; Wilder, 1950). The culmination of this work has resulted in a generalized theory of arousal, which more recently has proven to be a powerful explanatory tool for researchers attempting to account for psychological deficit in schizophrenia (e.g., Broen, 1968; Lang & Buss, 1965; Venables, 1964, 1966). However, a major criticism of generalized arousal theory is that the indices of arousal (electrocortical, neuromuscular, autonomic) show low, sometimes insignificant intercorrelations, especially when taken across individuals (Lacey, 1967). Thus an inference concerning arousal must be based on more than one response system.

The literature on the psychophysiology of schizophrenia presents much the same problem pertaining to measurement of arousal level. Moreover, the data seem most divergent when comparing indices within the autonomic nervous system. For example, in a recent review, Lang and Buss (1965) concluded that only the cardiovascular system has consistently differentiated normals and schizophrenics in terms of arousal level (chronic schizophrenics showing autonomic nervous system overarousal) whereas electrodermal activity has produced inconsistent results. Goldstein and Acker (1967) and Broen (1968) have come to similar conclusions concerning these autonomic nervous system indicators. In view of the present status of arousal theory, it would be most valuable to demonstrate that chronic schizophrenics are overaroused on an autonomic measure other than those which reflect cardiovascular functioning.

Because the electrodermal data appeared to be inconsistent with the general trend of psychophysiological research in schizophrenia (i.e., overarousal in these subjects), a more microscopic examination of these results seemed in order. Our review of the literature confirmed the previous conclusions that, in general, schizophrenics differ electrodermally from normals neither in terms of tonic levels nor with respect to phasic response (i.e., galvanic skin response) amplitudes. Exclusion of studies including patients receiving medication did not alter these conclusions. Since these studies have already been reviewed by others (Lang & Buss, 1965), they will not be discussed here. Two aspects of electrodermal activity did, however, appear to produce consistent results indicating greater arousal for schizophrenics, and those are reviewed below.

---

[1] The preparation of this manuscript was supported in part by Biomedical Sciences Support Grant FR-07035 from the National Institutes of Health. The money was made available in the form of a small grant for health-related research awarded by the Graduate College of the University of Iowa. The authors wish to thank Andrew L. Dickson for his critical reading of the manuscript.

[2] Requests for reprints should be sent to Richard A. Depue, Department of Psychology, Southern Illinois University, Carbondale, Illinois 62901.

Spontaneous Responses

In continuous recordings of electrodermal activity, nonspecific or spontaneous fluctuations which are uncorrelated in time with specific stimuli presented by the experimenter frequently occur. They are defined in the same manner as are phasic responses[3] except that they occur outside of the latency limits set for specific responses (usually 1–3 seconds after stimulus onset). There is considerable evidence that measures of frequency of responses are correlated with arousal level in normals (Burch & Greiner, 1960; Greiner & Burch, 1955; Silverman, Cohen, & Shmavonian, 1959; Surwillo & Quilter, 1965) and are independent of base-line measures (Lacey & Lacey, 1958), high frequencies being associated with increased arousal. Similarly, this measure has been used extensively as an index of arousal in psychiatric populations (Lader, 1971).

Of the investigators using spontaneous-fluctuation measurement, all have reported greater frequencies for schizophrenics except those employing patients on phenothiazines (Bernstein, 1964, 1970; Fenz & Velner, 1970). Thus, under conditions of minimal, nonsignal stimulation, Zahn (1964) and Zahn, Rosenthal, and Lawlor (1968) reported spontaneous fluctuation frequency in schizophrenics (both acute and chronic) to be three times that of normals, though this difference decreased during a simple reaction time task in another study. Moreover, in a ratio of specific to nonspecific responses analogous to a signal-to-noise ratio, normals had a reliably higher ratio than schizophrenics, indicating that the greatest percentage of the latter group's responses were to nonspecific stimulation.

Using similar stimulus conditions, Bernstein (1967) found both regressed and remitted patients to produce a greater number of spontaneous fluctuations than controls, while other researchers (Fowles, Watt, Maher, & Grinspoon, 1970) reported significantly greater than normal spontaneous fluctuation activity in good and poor premorbid patients under resting conditions, with poors exhibiting the greatest frequency.

Investigations utilizing psychologically stressful conditions revealed similar results. Thus, Syz (1926) and Syz and Kinder (1928) reported 21% and 8% more spontaneous fluctuation activity in paranoid and catatonic patients, respectively, than in controls during periods of listening to emotionally provoking words. No response to words was required. Furthermore, Syz reported data similar to Zahn's (1964; Zahn et al., 1968) mentioned earlier in that paranoids and catatonics revealed fewer responses to specific stimuli (12% and 29% fewer, respectively). Most striking here is the finding of similarity between paranoids and other schizophrenia subtypes in terms of spontaneous fluctuation frequency and signal-to-noise raito, since most other measures reveal divergence for these subjects. Similarly, Jurko, Jost, and Hill (1952), comparing early paranoids and normals, observed greater frequency of additional responses in patients during recovery periods following pictures depicting frustrating circumstances.

Finally, Stern, Surphlis, and Koff (1965) found that when tested periodically with tone stimulation (70 decibels) over a 5-week period, schizophrenics on phenothiazines, previously designated as having a good prognosis, significantly decreased in spontaneous fluctuation frequency while poor prognostic patients maintained their initial frequency. Though similar data were reported for verbal stimuli, the results were less dramatic than those for tones. Since the two groups did not differ prior to drug treatment in terms of spontaneous fluctuation frequency, the attractive notion that decreasing spontaneous fluctuation activity over time is indicative of improved health deserves further investigation. Moreover, the efficacy of frequency measures in differentiating schizophrenia subtypes when patients are on drug treatment is most interesting and also deserves further research. Unfortunately, the lack of a control group precluded comparisons with normal subjects.

---

[3] Phasic responses are usually defined as either a response of at least 300 ohms (this varies 300–500 ohms) deflection occurring within 1–3 seconds following specific stimulus onset (e.g., Bernstein, 1970; Zahn, Rosenthal, & Lawlor, 1968), a change score from basal level (e.g., Pishkin & Hershiser, 1963), or the largest responses in an individual's record (e.g., Goldstein & Acker, 1967).

A striking feature of the above studies is their demonstration that spontaneous fluctuation frequency measures differentiate patients and controls under a variety of stimulus conditions and not just those of minimal stimulation. However, some evidence suggests that conditions requiring overt responses on the part of schizophrenics are less efficacious in that both Fowles et al. (1970) and Zahn (1964) reported less difference in spontaneous fluctuation frequency between patients and controls during reaction time tasks than was found under conditions of rest (Fowles et al., 1970) or minimal, nonsignal stimulation (Zahn, 1964).

Habituation

Habituation to an experimental situation or to a series of stimuli may be quantified in two ways with electrodermal measurements. The more traditional way is by measuring the amplitude or frequency (i.e., probability) of the orienting response to a series of stimuli, habituation being defined as a decrease in response amplitude or probability. A less conventional method is to use measurements of skin conductance level or skin potential level with habituation being defined as a decline in these measurements from the beginning to the end of a specified time period. Habituation in this second sense can be used in the absence of specific stimulation, for example, rest periods. Since skin conductance level and skin potential level decline in the absence of sweat gland responses because of reabsorption of sweat in the duct (Fowles, 1972), the extent of this type of habituation is inversely related to the amount of spontaneous sweat gland activity during rest periods or to the combined effect of specific and spontaneous responses during a series of stimulus presentations. The greater spontaneous fluctuation activity found for schizophrenics would lead to a prediction of impeded habituation for these subjects relative to normals. This has been found to be the case.

The most extensive analysis of habituation in schizophrenia was undertaken by Zahn (1964) and Zahn et al. (1968) using tones and lights of low intensity. While analysis of basal levels in these studies revealed higher overall skin conductance level in chronic patients, this result was attributable to a failure to show a decrease in skin conductance level across trials in the patient group relative to normals, and similar trends were found for spontaneous fluctuation frequency over time. The results were similar for both tone and light stimulation, though more pronounced with light.

Turning to analysis of orienting responses to specific stimuli, Zahn et al. (1968) found similar results. In a subsample of patients matched with normals on mean and range of total orienting response frequency, schizophrenics showed invariant responding in terms of the number of specific orienting responses across trials, whereas normals steadily decreased in number. Upon closer inspection of the data, it was discovered that while normal subjects were characterized by gradual diminution of orienting response frequency, schizophrenics showed irregular stop-and-go responding. In light of the lower specific/nonspecific ratio found for patients, Zahn et al. (1968) suggested that this group's slower rate of habituation to specific stimuli was due in part to constant yet partial dishabituation resulting from attentional shifts to novel, nonspecific stimuli. Though this process would not preclude complete habituation, it would nevertheless impede its rate.

One additional measure of orienting response habituation confirmed the above results (Zahn, 1964; Zahn et al., 1968). With zero amplitude responses eliminated from analysis, both acute and chronic schizophrenics exhibited slower habituation of specific orienting response amplitude than controls across trials. A similar finding was reported for active schizophrenics in a study utilizing similar stimulus conditions (Venables, 1960), although differences did not hold when background stimulation was presented nor for withdrawn patients.

It is of interest that if the above analyses in Zahn et al.'s (1968) study had been computed on the first eight trials alone, the difference in habituation rates would not have been discovered. It was only after the first eight trials that normals began to decrease in response amplitude and frequency, whereas schizophrenics continued to respond in an invariant manner. Bernstein (1967) observed a similar divergence. While patients and controls showed equivalent responding initially, base-line and spontaneous fluctuation measures decreased in normals during the last 4 minutes of the 6-minute stimulation period, whereas both regressed and remitted schizophrenics revealed invariant basal reactivity and increasing spontaneous fluctuation frequency. Apparently in the Zahn and Bernstein studies, controls were exhibiting normal orienting response habituation while schizophrenics were not. Along the same lines, Israel (1966) reported slower habituation of skin conductance level for schizophrenics during exposure to mild photic stimulation, but no difference in overall skin conductance levels. Furthermore, the divergence over trials was greatest for a subsample of patients characterized as poor assimilators of novel stimulation (i.e., poor habituators). Similarly, Fowles et al. (1970) noted slower habituation of skin potential level in poor premorbids relative to goods and controls during an initial 10-minute rest period, while none of the groups differed in initial skin potential level. Poors also showed invariant spontaneous fluctuation responding during a subsequent reaction time task relative to normals.

Finally, Stern et al. (1965) again confirmed the value of using electrodermal activity as an indicator of improved health in patients. Upon admission, good and poor prognostic schizophrenics failed to differ in terms of habituation of orienting response reactivity to tones and a word-association task. However, 5 weeks later, poors were found to habituate less rapidly than goods, although the results were less pronounced with the word-association series.

In summary, then, the invariant responding by schizophrenics across time appears to be effective in differentiating normal and schizophrenic groups. In addition, a few studies (Fowles et al., 1970; Stern et al., 1965) suggest that patients with poor prognoses habituate less rapidly over time relative to good prognostic subjects. It appears that good and poor premorbids may not differ in habituation (and spontaneous fluctuation activity) upon admission, but that goods subsequently exhibit relatively normal patterns of responsivity more quickly than poors.

It is important to note that habituation effects are demonstrable with amplitude as well as frequency measures of electrodermal activity. One reason for this is that habituation is evaluated in terms of variations in amplitude within-subjects rather than in terms of the absolute amplitude of responses. The use of within-subject changes eliminates many of the problems of absolute measures, such as variations in sweat gland density, area of contact, etc., as most of these variables are constant for a given subject. As a result, intrasubject changes in response amplitude more accurately reflect changes in sweat gland activity and, therefore, sympathetic nervous system activity, than do absolute measures.

Summary and Conclusions

Measures of frequency of spontaneous responses yield consistently higher rates of response for schizophrenics whether measured during minimal stimulation (Bernstein, 1967; Fowles et al., 1970; Zahn, 1964; Zahn et al., 1968) or during more stressful stimulation (Jurko et al., 1952; Syz, 1926; Syz & Kinder, 1928). In contrast, the studies showing poor habituation of schizophrenics using either tonic levels or phasic response amplitudes have involved a minimum of stimulation (Bernstein, 1967; Fowles et al., 1970; Israel, 1966; Zahn, 1964; Zahn et al., 1968). Thus it appears that measures of intrasubject changes in tonic levels and phasic response amplitudes may be used most effectively at low levels of arousal, whereas frequency of response may be used at high or low levels of arousal. The reason for this has not been established, but it is likely that measures of level and response amplitude become less effective as arousal increases because of a ceiling effect or other limitation of the peripheral mechanism producing electrodermal phenomena.

Evidence has already been cited indicating that spontaneous fluctuations are indicative of arousal. The finding of poor habituation among anxiety neurotics (Lader & Wing, 1966) suggests that the poor habituation of schizophrenics is also a result of overarousal. Indeed, a large number of studies have shown that resistance to habituation and frequency of spontaneous responses are correlated (ranging from .40 to .74), suggesting that they may be viewed as alternate indicants of the same underlying dimension (Crider & Lunn, 1971). These results with schizophrenics, then, are consistent with those obtained with cardiovascular activity (Lang & Buss, 1965) and extend the findings of overarousal in schizophrenia to a second response system within the autonomic nervous system.

These conclusions apply to average results with chronic schizophrenics, and especially to nonparanoid chronic schizophrenics as defined by Venables (1964). It is also likely that there are other subgroups among chronic schizophrenics to which these conclusions do not apply. There is already suggestive evidence that active schizophrenics as assessed by the Venables' activity-withdrawal scale (Venables, 1957) may be underaroused. The activity-withdrawal dimension has been related to arousal, active patients being at the low end of the arousal continuum (Venables, 1960, 1963; Venables & Wing, 1962). Moreover, these patients have been found to exhibit inferior performance relative to normal and neurotic controls on an information-processing task under standard conditions but not when arousal was induced in all subjects by means of induced muscle tension.[4] Thus their performance was actually *improved* by increased arousal. Similarly, Venables (1960) found that active schizophrenics showed a reduction in the frequency of orienting responses with an increase in background stimulation. It seems unlikely, therefore, that active schizophrenics are overaroused. Further investigation of this group of patients would help to clarify the relationship between arousal and various types of behavior classified as schizophrenic.

There is no reason to believe that chronic schizophrenics are more aroused than are anxiety neurotics. Indeed a review of studies of anxiety neurotics would reveal results very similar to those reported for chronic schizophrenics. There must be an additional factor, probably some type of cognitive deficit, which contributes to schizophrenia in addition to overarousal. Patients who suffer from both chronic arousal and the cognitive deficit are likely to become chronic. This suggestion resembles the type of model proposed by Broen (1968), in which arousal interacts with a cognitive deficit (lowered maximum response strength ceiling) to produce the schizophrenic behavioral deficit. The cause of overarousal is a purely speculative matter at this point. The comparison with anxiety neurotics, however, suggests that it may be caused by an inhibitory deficit, as suggested by Malmo (1959).

REFERENCES

Bernstein, A. S. The galvanic skin response orienting reflex among chronic schizophrenics. *Psychonomic Science,* 1964, **1**, 391–392.

Bernstein, A. S. Electrodermal base level, tonic arousal, and adaptation in chronic schizophrenics. *Journal of Abnormal Psychology,* 1967, **72**, 221–232.

Bernstein, A. S. Phasic electrodermal orienting response in chronics schizophrenics: II. Response to auditory signals of varying intensity. *Journal of Abnormal Psychology,* 1970, **75**, 146–156.

Broen, W. E. *Schizophrenia: Research and theory.* New York: Academic Press, 1968.

Burch, N. R., & Greiner, T. H. A bioelectric scale of human alertness: Concurrent recording of the EEG and GSR. *Psychiatric Research Reports of the American Psychiatric Association,* 1960, No. 12, 183–193.

Crider, A., & Lunn, R. Electrodermal lability as a personality dimension. *Journal of Experimental Research in Personality,* 1971, **5**, 145–150.

Duffy, E. The concept of energy mobilization. *Psychological Review,* 1951, **58**, 30–40.

Fenz, W. D., & Velner, J. Physiological concomitants of behavioral indexes in schizophrenia. *Journal of Abnormal Psychology,* 1970, **76**, 27–35.

Fowles, D. C. Mechanisms of electrodermal activity. In R. F. Thompson & M. M. Patterson (Eds.), *Methods in physiological psychology.* New York: Academic Press, 1972.

Fowles, D. C., Watt, N. F., Maher, B. A., & Grinspoon, L. Autonomic arousal in good and poor premorbid schizophrenics. *British Journal of Social and Clinical Psychology,* 1970, **9**, 135–147.

Goldstein, M. J., & Acker, C. W. Psychophysiological reactions to films by chronic schizophrenics:

---

[4] R. A. Depue, and D. C. Fowles. Conceptual organization and arousal in active and withdrawn schizophrenics. Submitted for publication.

II. Individual differences in resting levels and reactivity. *Journal of Abnormal Psychology,* 1967, **72,** 23–29.

Greiner, T. H., & Burch, N. R. Response of human galvanic skin reflex to drugs that influence reticular formation of the brain stem. *Federation Proceedings,* 1955, **14,** 346.

Hebb, D. O. Drives and the C.N.S. (conceptual nervous system). *Psychological Review,* 1955, **62,** 243–254.

Israel, N. R. Individual differences in GSR orienting response and cognitive control. *Journal of Experimental Research in Personality,* 1966, **1,** 244–248.

Jurko, M., Jost, H., & Hill, T. S. Pathology of the energy system: An experimental clinical study of physiological adaptiveness capacities in a nonpatient, a psychoneurotic, and an early paranoid schizophrenic group. *Journal of Psychology,* 1952, **33,** 183–189.

Lacy, J. I. Somatic response patterning and stress: Some revisions of activation theory. In M. H. Appley & R. Trumbull (Eds.), *Psychological stress: Issues in research.* New York: Appleton-Century-Crofts, 1967.

Lacy, J. I., & Lacey, B. C. The relationship of resting autonomic activity to motor impulsivity. In, *The brain and human behavior.* (Research Publication of the Association for the Research of Nervous and Mental Disorders, Vol. 36) Baltimore: Williams & Wilkins, 1958.

Lader, M. H. Physiological measures in psychiatry. *Scientific Basis of Medicine Annual Reviews,* 1971, 279–304.

Lader, M. H., & Wing, L. *Physiological measures, sedative drugs, and morbid anxiety.* London: Oxford University Press, 1966.

Lang, P. J., & Buss, A. H. Psychological deficit in schizophrenia: II. Interference and activation. *Journal of Abnormal Psychology,* 1965, **70,** 77–106.

Lindsley, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

Malmo, R. B. Activation: A neurophysiological dimension. *Psychological Review,* 1959, **66,** 376–386.

Silverman, A. J., Cohen, S. I., & Shmavonian, B. M. Investigation of psychophysiological relationships with skin resistance measures. *Journal of Psychosomatic Research,* 1959, **4,** 65–87.

Stern, J. A., Surphlis, W., & Koff, E. Electrodermal responsiveness as related to psychiatric diagnosis and prognosis. *Psychophysiology,* 1965, **2,** 51–61.

Surwillo, W. W., & Quilter, R. E. The relation of frequency of spontaneous skin potential responses to vigilance and to age. *Psychophysiology,* 1965, **1,** 272–276.

Syz, H. C. Psychogalvanic studies in schizophrenia. *Archives of Neurology and Psychiatry,* 1926, **16,** 747–760.

Syz, H. C., & Kinder, E. F. Electrical skin resistance in normal and in psychotic subjects. *Archives of Neurology and Psychiatry,* 1928, **19,** 1026–1035.

Venables, P. H. A short scale for rating "activity-withdrawal" in schizophrenics. *Journal of Mental Science,* 1957, **103,** 197–199.

Venables, P. H. The effect of auditory and visual stimulation on the skin potential response of schizophrenics. *Brain,* 1960, **83,** 77–92.

Venables, P. H. Selectivity of attention, withdrawal, and cortical activation. *Archives of General Psychiatry,* 1963, **9,** 74–78.

Venables, P. H. Input dysfunction in schizophrenia. In B. Maher (Ed.), *Progress in experimental personality research.* Vol. 1. New York: Academic Press, 1964.

Venables, P. H. Psychophysiological aspects of schizophrenia. *British Journal of Medical Psychology,* 1966, **39,** 289–297.

Venables, P. H., & Wing, J. K. Level of arousal and the subclassification of schizophrenia. *Archives of General Psychiatry,* 1962, **7,** 114–119.

Wilder, J. The law of initial values. *Psychosomatic Medicine,* 1950, **12,** 392–401.

Zahn, T. P. Autonomic reactivity and behaviour in schizophrenia. *Psychiatric Research Reports,* 1964, **19,** 156–173.

Zahn, T. P., Rosenthal, D., & Lawlor, W. G. Electrodermal and heart rate orienting reactions in chronic schizophrenia. *Journal of Psychiatric Research,* 1968, **6,** 117–134.

(Received January 24, 1972)

Psychological Bulletin
1973, Vol. 79, No. 4, 239–242

THE DEVIL RIDES AGAIN:

CORRELATION AS AN INDEX OF FIT[1]

MICHAEL H. BIRNBAUM[2]

University of California, Los Angeles

Correlations between theoretical predictions and data can be higher for incorrect than for correct models, as illustrated by analyses of two sets of hypothetical data. This fact raises questions about the conclusions of recent studies that use correlation as an index of fit. Functional measurement provides a sounder basis for model evaluation by placing scaling in the context of model fitting and by testing deviations from prediction rather than concentrating an overall goodness of fit.

Although widely recognized as an "instrument of the devil" when used to infer causation from confounded data, the correlation coefficient is still employed as an index of the fit of theoretical models. There are two serious criticisms of this usage. First, incorrect models can correlate highly with data (Anderson, 1971; Yntema & Torgerson, 1961).

Second, based upon the assumption that the better model will correlate higher, investigators have recently used the correlation coefficient to compare the fit of rival models.[3] The point of the present paper is to warn that the devil may also be at work when correlation is used in this way: Predictions from an incorrect model sometimes correlate better with the data than predictions from the correct model.

LINEAR MODEL CORRELATES WITH MULTIPLICATIVE DATA

Table 1 illustrates one way this can happen. The matrix of hypothetical, errorless data might represent clinical ratings of the degree of neuroticism, Y, based upon two test scores, X_1 and X_2, with eight and three levels, respectively.

The usual *linear* model can be written as follows:

$$Y = a + bX_1 + cX_2, \qquad [1]$$

where Y is the predicted value for the data in the table, X_1 and X_2 are the a priori values of the two predictor variables (in this case, test scores), and a, b, and c are the linear coefficients.

The *multiplicative* model (Einhorn, 1970) can be written[4]

$$Y = aX^{b}_{1}X^{c}_{2}, \qquad [2]$$

where Y, X_1, and X_2 are defined as above, and a, b, and c are constants. Taking loga-

---

[1] Computing assistance was received from Campus Computing Network, University of California, Los Angeles. The author is grateful to Allen Parducci, Norman H. Anderson, Amos Tversky, Clairice T. Veit, and Andrew L. Comrey for their helpful comments on earlier versions of this paper.

[2] Requests for reprints should be sent to Michael H. Birnbaum, Department of Psychology, University of California, San Diego, P.O. Box 109, La Jolla, California 92037.

[3] For example, Einhorn (1970, 1971) and Goldberg (1971) have compared linear with "conjunctive" (multiplicative) models of judgment by correlating the data for each subject with the predictions from each of the rival models; when one model correlated higher for the great majority of the subjects, it was assumed to be a better representation of the process of human judgment. Einhorn (1971) obtained rankings of the attractiveness of hypothetical jobs, based on such cues as income and the opportunity to use special interests. The multiplicative model yielded higher correlations than the linear model for 32 of the 37 subjects. Goldberg (1971) found that the linear model yielded higher correlations than the multiplicative for each of 29 clinicians, who attempted to differentiate neurosis from psychosis on the basis of the Minnesota Multiphasic Personality Inventory profiles.

[4] Einhorn (1970) has defined Equation 2 as the "conjunctive model." Equation 2 captures some of the intuition behind the notion of a subjective "conjunctive" strategy; however, the multiplicative model (Equation 2) is compensatory and therefore should not be confused with the traditional definition of the conjunctive model (see Coombs, 1964).

TABLE 1

Hypothetical Multiplicative Data

A priori value of second cue	A priori value of first cue							
	1	2	3	4	5	6	7	8
1	15	18	21	24	27	30	33	36
2	20	22	24	26	28	30	32	34
3	25	26	27	28	29	30	31	32

TABLE 2

Hypothetical Additive Data

A priori value of second cue	A priori value of first cue				
	1	2	4	8	16
1	14	15	16	17	18
2	15	16	17	18	19
4	16	17	18	19	20
8	17	18	19	20	21
16	18	19	20	21	22

rithms of both sides of Equation 2 gives:

$$\log Y = \log a + b \log X_1 + c \log X_2, \quad [3]$$

which, like Equation 1, can be fitted by multiple linear regression with three constants.

The fit of Equation 1 to the data in Table 1 yielded a correlation of .930. Figure 1A plots these data against the best-fit predicted values in the usual way; the fit looks reasonably good. However, Figure 1B shows the same data plotted as a function of X_1, with a separate curve for each level of X_2. The nonparallelism of these curves indicates serious violation of the additive model of Equation 1. It should be clear that the data conform instead to a multiplicative model.

Einhorn's technique would fit his multiplicative model to the data by means of Equation 3, which yields a coefficient of only .899. This example shows that even when the data are perfectly multiplicative, comparison of correlation coefficients can lead to the erroneus conclusion that the linear model provides the better representation. The source of this difficulty lies in the assumption, implicit in the use of Equation 3, that the a priori values of X and Y are positive and known to a ratio scale. As shown below, the data of Table 1 fit Equation 2 perfectly when linear transformations are applied to the a priori scales.

Multiplicative Model Correlates with Additive Data

A second example shows that predictions from the multiplicative model can correlate better than those from the linear model, even when the linear model is more appropriate. The hypothetical data in Table 2 might be

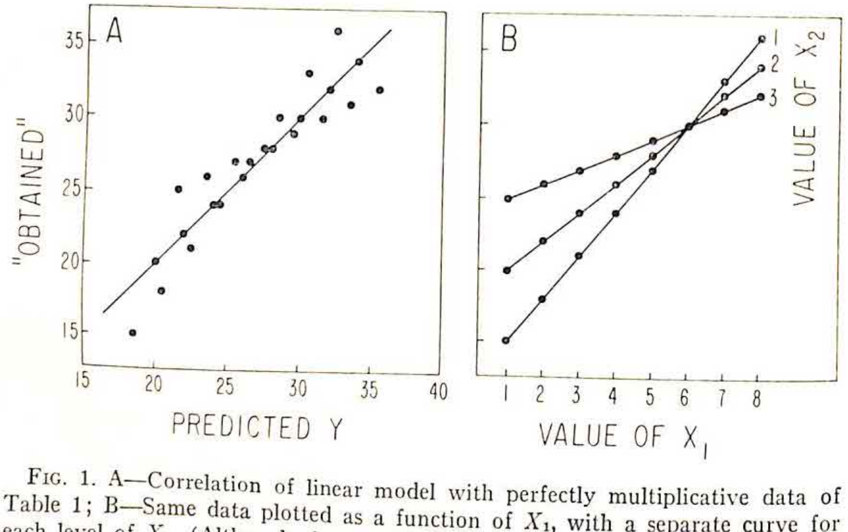


Fig. 1. A—Correlation of linear model with perfectly multiplicative data of Table 1; B—Same data plotted as a function of X_1, with a separate curve for each level of X_2. (Although the intersecting linear functions in B indicate that the data are perfectly multiplicative [with a functional zero point where the curves cross], Equation 3 provides an inferior fit because it requires the curves to intersect at the zero point of the a priori X and Y scales.)

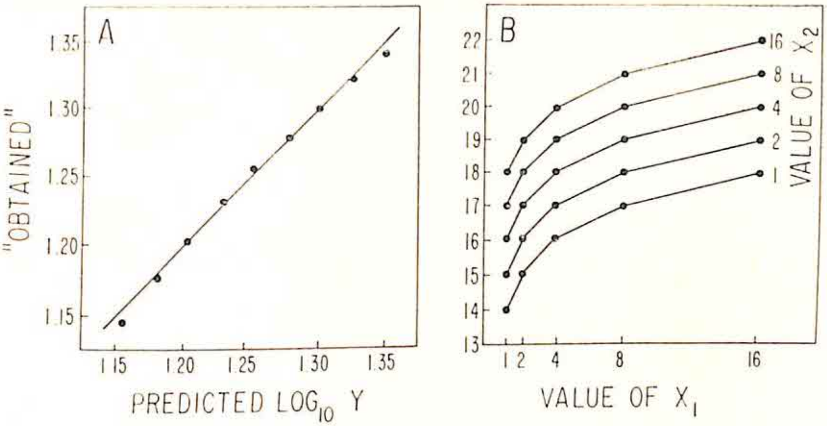


FIG. 2. A—Correlation of multiplicative model with perfectly additive data of Table 2. (Graph points represent varying numbers of data points.) B—Same data plotted as a function of X_1, with a separate curve for each level of X_2. (Although the parallelism of the curves in B indicates perfect additivity, the linear model provides an inferior fit because it requires the curves to be straight-line functions of the a priori values of X_1 and X_2.)

ratings, Y, of the attractiveness of various jobs based on two cues, X_1 and X_2.

Fitting the linear model to these data yields a correlation of only .933 compared with .998 for the multiplicative model. It can be inferred from Figure 2A, which plots the fit of Equation 3 to the data, that the use of rank-order correlation (Einhorn, 1970, 1971) would have "improved" on the already excellent fit of the wrong model. The multiplicative model is the wrong model because the data are perfectly additive, as shown by the parallelism of the curves of Figure 2B. The investigator using correlation as an index of fit might erroneously conclude that the data reveal a "conjunctive" judgmental strategy unless the data were plotted as in Figure 2B. It is only because the linear regression model requires the data to plot as linear functions of the a priori values of X_1 and X_2 that it yields a lower correlation.

FUNCTIONAL MEASUREMENT PROVIDES BETTER TEST

Functional measurement (Anderson, 1970, 1971) correctly diagnoses the same hypothetical data. The linear model is additive and therefore predicts a zero interaction between the cues. Equation 2 is multiplicative and therefore predicts an interaction which should be located entirely in the bilinear component. These interactions can be tested graphically, as in Figure 1B and 2B, and statistically, using the analysis of variance.

The intersecting linear curves of Figure 1B are indicative of a multiplicative model with a functional zero at the point where the curves intersect. When Y is replaced by $Y - 30$, X_1 is replaced by $X_1 - 6$, X_2 is replaced by $4 - X_2$, and $a = b = c = 1$, Equation 2 fits the data of Figure 1B perfectly.

The parallelism of the curves of Figure 2B indicates that there is no interaction between cues, in agreement with an additive model. In the case of additive models, the marginal means estimate the functional scale values for the cues. Equation 1 fits these data perfectly when the marginal means of Table 2 are substituted for the a priori values of X.

COMMENTS AND CONCLUSIONS

These two examples demonstrate that even with factorial designs and errorless, interval data, the correlation coefficient can be an inappropriate instrument for comparing the fit of models with equal numbers of parameters. In real-life applications, correlations may be even more misleading.

Correlations can be diabolical when factorial designs are not employed. The reader can easily select those cue combinations from

Figure 1B and 2B that would improve the fit of either model. Furthermore, with certain research designs, graphic tests like those shown in Figures 1B and 2B become impossible. Hence, "representative" or improper "contrived" designs can exaggerate the fit of a seriously defective model and also can make it impossible to evaluate.

The data need not be particularly unusual for the correlations to be so misleading. In Table 2, the functional scale values (marginal means) are linearly related to the logarithms of the a priori stimulus values. That is often the case for category ratings of psychophysical stimuli or stimuli presented in numerical form, such as incomes. Consequently, there is a very real danger that the multiplicative model would yield higher correlations than the linear in many applications, even though the data were additive.[5]

In contrast to correlational procedures, functional measurement (Anderson, 1970, 1971) requires neither a priori values for the stimuli, nor ratio scales for the responses.[6] These procedures estimate best-fit values for the stimuli directly from the data, using these estimates as the basis for statistical tests of the discrepancies. Functional measurement typically employs factorial designs, analysis of variance, and graphical inspection of the crucial features of the data. These advantages of functional measurement over the correlational approach are illustrated by the sample data in the present paper.

---

[5] Goldberg (1971) has proposed that some of these problems would be avoided if a greater number of models were compared in any correlational contest of fit. For example, Goldberg's logarithmic model, which replaces X with log X in Equation 1, would have correlated perfectly with the data of Table 2. However, none of his five models would have correctly diagnosed the data of Table 1. The number of possible correlation-regression models is impractically large, since each arbitrary rescaling of either the a priori stimulus values or the overt responses is treated as a different model. It should be emphasized that the usual applications of multiple regression (as in the present examples) are *not* equivalent to the analysis of variance. Equivalent techniques (e.g., Cohen, 1968) are seldom employed in this type of analysis because they would be cumbersome to apply.

[6] Functional measurement also provides for monotonic rescaling of the data under certain circumstances although no nonlinear transformation was called for in these examples. When the dependent variable is considered to be only an ordinal measure of the psychological attribute, models can be compared by examination of critical ordinal properties of the data (Krantz & Tversky, 1971). For example, the crossover interaction in Figure 1B is ordinally inconsistent with an additive representation.

REFERENCES

Anderson, N. H. Functional measurement and psychophysical judgment. *Psychological Review,* 1970, 77, 153–170.

Anderson, N. H. Integration theory and attitude change. *Psychological Review,* 1971, 78, 171–206.

Cohen, J. Multiple regression as a general data-analytic system. *Psychological Bulletin,* 1968, 70, 426–443.

Coombs, C. H. *A theory of data.* New York: Wiley, 1964.

Einhorn, H. J. Use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin,* 1970, 73, 221–230.

Einhorn, H. J. Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organizational Behavior and Human Performance,* 1971, 6, 1–27.

Goldberg, L. R. Five models of clinical judgment: An empirical comparison between linear and nonlinear representations of the human inference process. *Organizational Behavior and Human Performance,* 1971, 6, 458–479.

Krantz, D. H., & Tversky, A. Conjoint measurement analysis of composition rules in psychology. *Psychological Review,* 1971, 78, 151–169.

Yntema, D. B., & Torgerson, W. S. Man-computer cooperation in decisions requiring common sense. *IRE Transactions on Human Factors in Electronics,* 1961, HFE-2, 20–26.

(Received January 26, 1972)

Psychological Bulletin
1973, Vol. 79, No. 4, 243-251

SHOCK-MOTIVATED AVOIDANCE AND THE ANALYSIS OF BEHAVIOR[1]

DAVID S. OLTON[2]

Johns Hopkins University

Avoidance tasks are distinguished by their spatial characteristics and by whether response initiation or response suppression is required. Justification is given for a resulting classification system composed of passive avoidance, one-way active avoidance, two-way active avoidance, etc. This system is then used to discuss (*a*) the interrelationship of motivation and performance, (*b*) cue utilization and species differences in cue utilization, (*c*) behavioral changes following brain lesions, (*d*) comparison of behavioral roles of homologous brain structures, (*e*) theoretical and conceptual models of central nervous system functioning. The system is useful in resolving apparent discrepancies, summarizing relatively large bodies of data, and distinguishing among alternative explanations of results.

The proliferation of shock-motivated testing procedures has produced many apparently discrepant results. Resolution of these discrepancies, based on consideration of spatial and response characteristics, has been offered, but only within a specific theoretical (McCleary, 1966), experimental (Olton & Isaacson, 1968a; Theios, Lynch, & Lowe, 1966), or a neurophysiological (Olton & Isaacson, 1968b) framework. Consequently, the utility of this approach has not been widely recognized. This paper reviews the spatial and response distinctions mentioned above, emphasizing their general applicability and indicating their usefulness in summarizing large bodies of data and in distinguishing among alternative theoretical explanations of results.

The first major distinction to be drawn among avoidance tasks is that between active and passive avoidance (McCleary, 1961; Mowrer, 1960). In the active avoidance paradigm, the animal must emit a particular response to escape or avoid a noxious stimulus, while in the passive avoidance paradigm the animal must *not* emit a particular response. In principle, the appropriate time for the response may or may not be signaled by a discriminative stimulus. In practice, active avoidance tasks almost always utilize discriminative stimuli while passive avoidance tasks do not and unless otherwise noted, this convention is followed throughout the discussion that follows.

In the typical active avoidance task, a conditioned stimulus (CS) is presented in the presence of which the animal must emit some well-defined low-probability response such as pressing a bar, jumping over a barrier, climbing a pole, etc., in order to avoid shock. Measures include the number of trials required to perform the avoidance response reliably and the latency of those responses. In the typical passive avoidance task, conditions are arranged so that the animal has a high probability of emitting a response. This may result from training or from a species-typical response such as stepping off a small platform, entering a dark compartment, leaving a shaking bar, etc. At some time arbitrarily chosen by the experimenter, this response is punished. Passive avoidance measures include a decrease in the probability and an increase in the latency of response.

The terms "active" and "passive" are used only to define the procedures just described; no other connotations are intended. Also, these terms refer only to the experimenter-defined response. During any avoidance training a variety of behaviors are observed. In passive avoidance the animal rarely does nothing as the term might imply; rather, it engages in a series of approach–avoidance and displacement activities. Similarly, in active avoidance the animal does not just emit the correct response but suppresses some (such as freezing) and engages in others (such as exploration). In short, the terms "active avoidance" and "passive avoidance" do not completely describe the animal's behavior. Nonetheless, it is difficult to find more appropriate terms, and this paper will continue to use the active–passive terminology.

---

[1] The author received support from a Johns Hopkins Faculty Grant (P59-6194) and from a National Institute of Mental Health Research Grant (MH 19124-01) during the writing of this manuscript. Comments from R. L. Isaacson and E. E. Krieckhaus upon earlier versions of the manuscript are gratefully acknowledged.

[2] Requests for reprints should be sent to David S. Olton, Psychology Department, Johns Hopkins University, Baltimore, Maryland 21218.

A further division of active avoidance tasks into one-way active avoidance and two-way active avoidance completes the classification system. In one-way active avoidance the animal responds in only one direction, leaving a consistently noxious location and entering a consistently safe location. In two-way active avoidance, there are no consistently safe and noxious locations. The animal must learn to respond in two directions, leaving one location and entering the other on one trial, and leaving the latter and entering the former on the next trial.[3]

The one-way, two-way classification requires that an avoidance task have exactly two locations; not all avoidance tasks do. There may be only one location as in a bar-press avoidance task, or there may be several locations, as in a sequential avoidance task. Although both types of tasks may be analyzed in terms of one-way and two-way components (see p. 245), neither fits into the strict categorization given above and consequently is place into an "other" category.

MOTIVATION AND PERFORMANCE: STIMULUS INTENSITY

In the present context, the general problem concerning the interrelationship between intensity of motivation and efficiency of learning or performance becomes: How does variation of shock intensity affect the rat's avoidance behavior? In both one-way active avoidance and passive avoidance, the rate of acquisition increases or does not change as shock intensity is raised from .5 milliampere to 3.0 milliamperes (Kurtz & Pearl, 1960; Moyer & Korn, 1966; Theios et al., 1966). In two-way active avoidance, just the opposite is true. With shock intensities greater than .5 milliampere, the rate of acquisition of two-way active avoidance steadily declines as shock intensity rises (Beatty & Beatty, 1970; Davis, Mook, & Brush, 1958; Johnson & Church, 1965; Kurtz & Shafer, 1967; Levine, 1966; Martin, Powell, & Kamano, 1967; McAllister & McAllister, 1971; Moyer & Korn, 1964; Overmier & Seligman, 1967; Theios et al., 1966).

An appropriate post hoc explanation for the differences just described lies in the conflict characteristics of the two tasks. The one-way task is conflict free. One location is always noxious, the other is always neutral. Raising shock intensity ought to increase the aversive character of the former while leaving the neutral character of the latter unaffected. The animal should be more likely to move from the shocked to the unshocked location. The two-way task is not conflict free. In order to avoid shock on any given trial, the animal must return to a location in which it was shocked on the preceding trial. The resulting outcome is an avoidance–avoidance conflict. Raising the shock intensity ought to increase the conflict, a procedure which often produces maladaptive stereotyped behaviors which interfere with the correct response. Other analyses, in more detail, have been presented by McAllister and McAllister (1971) and McCleary (1966).

At extremely high intensities of shock, the monotonic interrelationship between shock intensity and rate of acquisition may break down. Dogs tested in two-way active avoidance with "subtetanizing" (5–10 milliamperes) levels of shock (Solomon & Wynne, 1953) attained criterion in very few trials. Although there is no comparable one-way avoidance experiment, an inversion of the monotonic trend might be found here also.

---

[3] It is suggested that the term "shuttle box" be discontinued because of the confusion between "shuttle box avoidance" and "shuttle box apparatus." The former refers to a two-way procedure while the latter refers to a two-compartment apparatus in which either one-way or two-way active avoidance may be conducted.

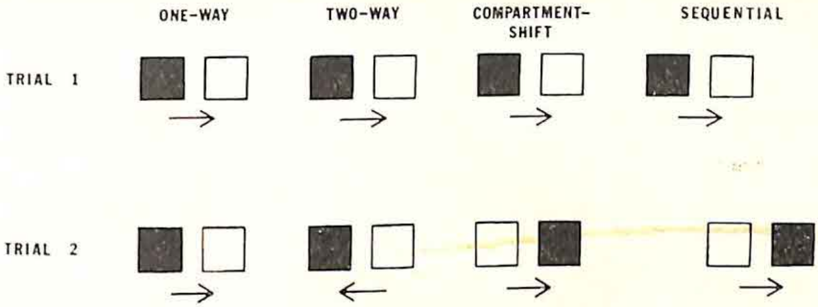


FIG. 1. Schematic representation of the four avoidance procedures discussed in Olton and Isaacson (1968a). (The arrows indicate direction of movement and the shading differentiates the two compartments. Note that in the sequential avoidance procedure the white compartment, containing the animal, is not moved during the intertrial interval.)

CUE UTILIZATION AND ATTENTION

Preferential reaction to certain stimulus modalities, and species differences in this characteristic are two other problems which can be dealt with by consideration of one-way and two-way active avoidance. Up to this point, no attempt has been made to indicate the characteristics distinguishing one location from another. This issue is important because (*a*) it provides information about cue preferences of different species and (*b*) it determines the classification of a particular procedure as either one-way or two-way active avoidance.

One study dealing directly with this issue measured performance in four different active avoidance tasks (Olton & Isaacson, 1968a). One-way and two-way procedures were conducted as usual with the two compartments remaining in their respective spatial locations. Next, the cues for spatial location and compartment identification were opposed. In the compartment-shift avoidance procedure (McCleary, 1966) the compartments (one of which contained the animal) were interchanged at the end of the intertrial interval which resulted in one-way active avoidance with respect to spatial location but two-way active avoidance with respect to compartments. In the sequential avoidance procedure, the compartment just electrified was moved ahead of the compartment containing the animal. This task was again two-way with respect to compartmental cues, but differed from both the one-way and two-way tasks with respect to spatial cues. Figure 1 illustrates these procedures.

Results indicated that performance in tasks spatially equivalent (compartment shift and one way) was similar, while performance in tasks spatially distinct (compartment shift, sequential, and two way) was diffent. Although compartmental cues are not irrelevant in avoidance performance (Denny, Koons, & Mason, 1959), the data suggest that rats preferentially associate shock with a given location rather than with coextensive visual and auditory stimuli.

Are other species as dependent upon spatial cues in the performance of avoidance tasks? For cats, the number of trials to criterion in compartment-shift avoidance is almost as large as that observed in two-way active avoidance (Lubar & Perachio, 1965; McCleary, 1966). These data suggest that the cat treats the compartment-shift task as different from one-way active avoidance and similar to two-way active avoidance. A possible complication lies in the method used to move the compartments (Kriekhaus, 1965) but other experiments (Baum & Bobrow, 1966) suggests that this may not be a consideration.

Additional data come from a comparison of the normal two-way active avoidance procedure with ones in which the location of the conditioned stimulus (CS) is changed from trial to trial so that the task is two way with respect to spatial location and compartment cue but one way with respect to the CS. The performance of cats is substantially affected by having an auditory CS presented in this one-way fashion (McAdam, 1964) while the performance of rats is not (Whittleton, Kostanek, & Sawrey, 1965). Again the data suggest that cats do not preferentially attend to cues regarding spatial location. Monkeys probably also attend to auditory and visual cues in such a procedure, rather than to spatial ones (Krieckhaus, 1967).

In summary, there are a variety of cues attended to during avoidance training (cf. Rozin and Kalat, 1971) and these vary from species to species. More direct tests concerning the importance of various sensory modalities in avoidance learning in animals other than the rat would be helpful for understanding the behavior of the species in question, and for assistance in categorizing a particular avoidance procedure as either one-way or two-way active avoidance.

BRAIN LESIONS AND AVOIDANCE BEHAVIOR

Early studies (Papez, 1937) correctly suggested that the neuroanatomical bases of emotional behavior were in subcortical, especially limbic structures, but it was only with subsequent testing in shock-motivated avoidance procedures that finer distinctions could be made. One well-known study (McCleary, 1961) demonstrated that septal lesions impaired passive avoidance but not active avoidance, while cingulate lesions impaired active avoidance but not passive avoidance. Other experiments have corroborated these distinctions for the two brain areas mentioned and have extended them to many other subcortical structures. As will be indicated, these results provide additional support for the classification of avoidance tasks presented here and allow finer distinctions to be drawn among the behavioral roles of the neuroanatomical substrates of emotion.

Much of the material relevant to this issue has been covered in detail in recent review articles. Consequently, in the ensuing discussion reference is made to review articles rather than to original sources whenever possible. Except where indicated, only data from rats is considered.

There is little systematic information concerning the results of neocortical lesions on avoidance behavior. With small lesions of the posterior cortex, rats appear to be slightly impaired in the acquisition and reacquisition of both one-way and two-way active avoidance tasks, exhibiting longer response latencies (Olton & Isaacson, 1968b; Saavedra, Pinto Hamuy, & Oberti, 1965). Passive avoidance is probably unaffected (Isaacson, Olton, Bauer & Swart, 1966; Isaacson & Wickelgren, 1962; Kimble, 1963; Kimble, Kirkby, & Stein, 1966; Olton & Isaacson, 1968b). In the cat, lesions of the lateral gyrus, posterolateral gyrus, and occipital cortex impair acquisition of two-way active avoidance (Lubar & Middaugh, 1971; Lubar, Perachio, & Kavanagh, 1966; Lubar, Schastal, & Perachio, 1967).

Lesions in the dorsomedial nucleus of the thalamus produce an impairment in acquisition of both one-way and two-way active avoidance, but have no effect upon acquisition of passive avoidance. Postoperative retention in the absence of shock is unaffected (Olton & Isaacson, 1967). This suggests that the behavioral change is in response to the exteroceptive somatic stimulus (pain) rather than to an internal representation (fear). For a review of this literature, see Vanderwolf (1971).

Other thalamic nuclei are also involved in avoidance behavior. Small lesions of the midline nuclei impair acquisition of two-way active avoidance (Bohus & de Weid, 1967) and reacquisition of one-way active avoidance, Thompson, 1963) but escape behavior remains unchanged. After destruction of the parafascicularis nucleus, rats are deficient in acquistion of one-way avoidance, but unimpaired in acquisition of two-way active avoidance (Bohus & de Weid, 1967). Lesions of the anterior thalamic nuclei, part of the original Papez "circuit," interfere with acquisition of passive avoidance (Gerbrandt, 1965) but not with the retention of one-way active avoidance (Rich & Thompson, 1965). Lesions of the lateral nucleus or the ventral nucleus do not alter avoidance behavior (Olton & Isaacson, 1967; Thompson, 1963).

Hippocampal lesions in rats enhance acquisition of two-way active avoidance and either slightly impair or do not affect acquisition of one-way active avoidance. An impairment in passive avoidance may or may not be found depending upon the characteristics of the task (Nadel, 1968). For reviews of this literature, see Douglas (1967) and Olton and Isaacson (1968b).

Septal lesions in rats produce behavioral changes similar to those found after hippocampal lesions. Acquisition of passive and one-way active avoidance is impaired while acquisition of two-way active avoidance is enhanced. In cats, septal lesions impair acquisition of passive avoidance and enhance acquisition of compartment-shift avoidance. The changes in passive avoidance are independent of altered regulatory function and the changes in active avoidance are independent of the septal hyperreactivity snydrome. For reviews, see McCleary (1966) and Thomas, Hostetter, & Barker (1968).

The effects of amygdaloid lesions in rats upon avoidance behavior are not clear, partly due to the lack of agreement upon appropriate subdivisions of the amygdaloid nuclei. There are no reports of facilitation of avoidance behavior after lesion or stimulation of the amygdala, while reports of deficits are common. Additional data are needed before a definite statement can be made about the specific pattern of performance changes exhibited on different avoidance tasks (see Goddard, 1964).

Caudate lesions in rats impair acquisition of one-way active avoidance and of trained, but not untrained, passive avoidance (Kirkby & Kimble, 1968). In cats, caudate lesions impair passive avoidance (Fox, Kimble, & Lickey, 1964) and two-way active avoidance (Thompson, 1959; Thompson & Mettler, 1963).

Mammillothalamic tract lesions have little, if any, effect upon the acquisition of either

	Passive Avoidance	One-way Active Avoidance		Two-way Active Avoidance	
		Acq.	Ret.	Acq.	Ret.
NEOCORTEX	→	↘	↘	↘	↘
THALAMUS					
Anterior nucleus	↓				
Dorsomedial nucleus	→	↓	↓	↓	↓
Midline nuclei	↓	⇣	⇣	↓	⇣
HIPPOCAMPUS	↓	↘	↘	↑	→
SEPTUM	↓	↘	↘	↑	⇢
AMYGDALA	⇢	↘	⇢		⇢
CAUDATE NUCLEUS	↓	⇣		⇢	⇢
MAMMILLOTHALAMIC TRACT	→	→	⇢	→	↓

FIG. 2. Avoidance performance after brain lesions. (The arrows indicate: greatly facilitated performance ↑, slightly facilitated performance ↘, no change in performance →, slightly impaired performance ↘, greatly impaired performance ↓. A solid arrow indicates consistency in the data while a dashed arrow indicates confusion. No arrow indicates the absence of relevant data.)

passive avoidance, one-way active avoidance, or two-way active avoidance. Acquisition of more complex avoidance tasks, as well as retention of two-way active avoidance and of discriminated avoidance is impaired. For reviews, see Krieckhaus (1967).

Figure 2 summarizes these data, illustrating what is known about the performance of animals with particular brain lesions, and what has not yet been learned, either because there are no experimental data or because the data are conflicting.

Two conclusions may be drawn from these experiments, First, the behavioral effects of a particular central nervous system lesion vary depending upon the type of avoidance task. Second, the pattern of performance across all three types of avoidance tasks varies with the site of the lesion. These results indicate that the neuroanatomical mechanisms used during avoidance training differ from one type of avoidance task to another. These conclusions provide neuroanatomical support for the classification system proposed here and allow further distinctions to be drawn among the behavioral roles of different subcortical structures thought to be important in the elaboration of emotions.

Although further subdivision of the brain structures summarized is probably appropriate (Grant & Jarrard, 1968; Van Hosen, MacDougal, & Mitchell, 1969), there is not yet sufficient information about animals with more discrete lesions to include them in Figure 2. As more data become available, the same type of analysis applied to the larger structures should also be of value for investigation of their smaller subunits.

Comparative Studies

To determine whether or not homologous brain structures have equivalent behavioral functions, cross-species comparisons and correct classification of avoidance tasks are required. This procedure is made difficult because a given objective procedure may not represent the same kind of task for different species. Rats treat compartment-shift avoidance as if it were a one-way task; cats may or may not (see p. 246). Septal lesions in rats impair one-way active avoidance; septal lesions in cats do not affect compartment-shift avoidance. If compartment-shift avoidance for cats is equivalent to the one-way procedure for rats, then the spetal nuclei in these species do not have equivalent behavioral

	PASSIVE AVOIDANCE	ONE-WAY ACTIVE AVOIDANCE	TWO-WAY ACTIVE AVOIDANCE
INCREASED ACTIVITY	?	↑	↑
INCREASED "FEAR"	↑	↑	↓
DECREASED "FEAR"	↓	↓	↑
DECREASED RESPONSE INHIBITION	↓	↑	↑
DECREASED RESPONSE FACILITATION	→	↓	↓

FIG. 3. Some psychological hypotheses and the patterns of performance that would be expected to accompany them. (The arrows indicate: enhanced performance ↑, no change in performance →, and impaired performance ↓.)

functions. If, on the other hand, the compartment-shift procedure for cats is a type of hybrid between one-way and two-way avoidance, these homologous brain structures may have equivalent functions. Similar problems arise in consideration of cross-species comparisons after lesions of the mammillothalamic tract (Krieckhaus, 1967) and the hypothalamus (Ingram, 1958; McAdam & Kaelber, 1966), and further experiments are necessary to settle the issue.

THEORETICAL AND CONCEPTUAL MODELS

The avoidance analysis presented here is useful in summarizing data from experiments using brain lesions as well as offering suggestions as to the psychological processes that might have been altered as a result of the lesion. Some hypotheses and related performance patterns are presented in Figure 3. The rationale for most of the patterns is clear. Increased activity ought to facilitate acquisition of both types of active avoidance because it should decrease the probability of maladaptive freezing responses; passive avoidance would be affected only if the activity was goal directed. Changes in response facilitation and response inhibition, used in conceptual rather than descriptive context, ought to produce the patterns indicated.

Two relationships that might not be obvious are: increased fear, two-way active avoidance impaired, and decreased fear, two-way active avoidance enhanced. Evidence has been presented (see p. 244) that increasing shock intensity decreases the rate of acquisition of two-way active avoidance. If fear is positively correlated with shock intensity, increasing fear ought to retard acquisition of two-way active avoidance while decreasing fear ought to accelerate it.

Other hypotheses and patterns of performance are, of course, possible. Figure 3 is intended only to illustrate the advantage of the analysis proposed here. By equating all experimental parameters except the manner in which the shock is applied, undesirable artifacts may be eliminated. Both for the generation of new hypotheses and for the rigorous testing of already proposed hypotheses, the pattern of performance on these three tasks offers a rapid and reliable procedure.

When these theoretical patterns of performance are compared with those actually observed after brain lesions, it is apparent that changes in activity are probably not the cause of avoidance changes following any brain lesions, although they may be observed as independent (Douglas, 1967) or dependent (Krieckhaus, 1967) effects. Increased fear seems unacceptable also. Decreased fear is compatible with the results of septal lesions (Kenyon & Krieckhaus, 1965; Lubar, 1964) and hippocampal lesions (Olton & Isaacson, 1969). Decreased (voluntary) response facilitation is compatible with the results of dorsomedial thalamic lesions (Vanderwolf, 1971), while a change in the ability to shift from stereotyped to novel responses has been proposed for the results observed after mammillothalamic tract lesions (Krieckhaus, 1967). Pending appropriate classification of the compartment-shift task for cats, either decreased response inhibition or decreased fear might be appropriate explanations for the behavioral changes observed after septal lesions.

SUMMARY AND CONCLUSIONS

Justification has been offered for classification of avoidance tasks into passive avoidance, one-way active avoidance, and two-way active avoidance. This analysis has been applied to five different areas, and the following conclusions drawn: (*a*) the amount of conflict present in a task influences the relationship between intensity of motivation and rate of learning; (b) cues concerning spatial location in avoidance behavior are dominant in the rat, but probably not in the cat or monkey; (*c*) the functional role of subcortical structures thought to be involved in emotional behavior can be reliably differentiated; (d) appropriate classification of avoidance tasks is mandatory in determining whether homologous brain structures have equivalent behavioral functions; (*e*) theoretical and conceptual models of central nervous system functioning can be verified by comparison of performance on these three tasks.

Criticism has been made of experiments utilizing shock avoidance procedures because these behaviors are "unnatural" for animals

(Bolles, 1970). The evidence presented here suggests that this is not the case. Certainly an animal in the wild rarely, if ever, experiences electric shock, set intertrial intervals, etc. Nonetheless, the generality of the approach proposed here indicates that appropriate shock motivated procedures utilize brain systems naturally designed to cope with aversive events. The distinctions drawn here among avoidance tasks appear to be significant both for the animal's behavior and for its central nervous system, suggesting that these procedures are appropriate for investigation of aversively motivated behaviors and for understanding the central nervous system mechanisms that underly them.

REFERENCES

Baum, M., & Bobrow, S. A. An automated analogue of the one-way Miller-type avoidance box. *Psychonomic Science*, 1966, **5**, 361–362.

Beatty, W. W., & Beatty, P. A. Hormonal determinants of sex differences in avoidance behavior and reactivity to electric shock in the rat. *Journal of Comparative and Physiological Psychology*, 1970, **73**, 446–455.

Bohus, B., & de Weid, D. Avoidance and escape behavior following medial thalamic lesions in rats. *Journal of Comparative and Physiological Psychology*, 1967, **64**, 26–29.

Bolles, R. C. Species-specific defense reactions and avoidance learning. *Psychological Review*, 1970, **77**, 32–48.

Davis, C. G., Mook, D. G., & Brush, F. R. Acquisition of avoidance behavior as a function of shock intensity using a trace conditioning procedure. *American Psychologist*, 1958, **13**, 419. (Abstract)

Denny, M. R., Koons, P. B., & Mason, J. E. Extinction of avoidance as a function of escape situation. *Journal of Comparative and Physiological Psychology*, 1959, **52**, 212–215.

Douglas, R. J. The hippocampus and behavior. *Psychological Bulletin*, 1967, **67**, 416–442.

Fox, S., Kimble, D. P., & Lickey, M. E. Comparison of caudate nucleus and septal area lesions on two types of active avoidance behavior. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 380–386.

Gerbrandt, G. K. The effects of anteromedial and dorsomedial thalamic lesions on passive avoidance and activity. *Psychonomic Science*, 1965, **2**, 39–40.

Goddard, G. V. Functions of the amygdala. *Psychological Bulletin*, 1964, **62**, 89–109.

Grant, L. D., & Jarrard, L. E. Functional dissociation within hippocampus. *Brain Research*, 1968, **10**, 392–401.

Ingram, W. R. Modification of learning by lesions and stimulation in diencephalon and related structures. In H. H. Jasper (Ed.), *Reticular formation of the brain*. Boston: Little, Brown, & Company, 1958.

Isaacson, R. L., Olton, D. S., Bauer, B., & Swart, P. The effect of training trials on passive avoidance deficits in the hippocampectomized rat. *Psychonomic Science*, 1965, **3**, 497–499.

Isaacson, R. L., & Wickelgren, W. O. Hippocampal ablation and passive avoidance. *Science*, 1962, **138**, 1104–1106.

Johnson, J. L., & Church, R. M. Effects of shock intensity in nondiscriminative avoidance learning of rats in a shuttlebox. *Psychonomic Science*, 1965, **3**, 497–499.

Kenyon, J., & Krieckhaus, E. E. Decrements in one-way avoidance learning following septal lesions in rats. *Psychonomic Science*, 1965, **3**, 113–114.

Kimble, D. P. The effects of bilateral lesions in rats. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 273–283.

Kimble, D. P., Kirkby, R. J., & Stein, D. G. Response perserveration of passive avoidance deficits in hippocampectomized rats. *Journal of Comparative and Physiological Psychology*, 1966, **61**, 141–143.

Kirkby, R. J., & Kimble, D. P. Avoidance and escape behavior following striated lesions in the rat. *Experimental Neurology*, 1968, **20**, 215–227.

Krieckhaus, E. E. Reply to Zucker. *Psychonomic Science*, 1965, **3**, 310.

Krieckhaus, E. E. The mammillary bodies: Their function and anatomical connections. *Acta Biologica Experimentia* (Warsaw), 1967, **27**, 319–337.

Kurtz, H. J., & Pearl, J. The effects of prior fear experiences on acquired-drive learning. *Journal of Comparative and Physiological Psychology*, 1960, **53**, 201–206.

Kurtz, P. S., & Shafer, J. N. The interaction of UCS and intertrial interval in avoidance learning. *Psychonomic Science*, 1966, **5**, 409–410.

Levine, S. UCS intensity and avoidance learning. *Journal of Experimental Psychology*, 1966, **71**, 163–164.

Lubar, J. F. Effect of medial cortical lesions on the avoidance behavior of the cat. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 38–46.

Lubar, J. F., & Middaugh, L. D. Avoidance conditioning in cats with striate and auditory cortex lesions and contrasting effects of peripheral blindness. *Journal of Comparative and Physiological Psychology*, 1971, **76**, 26–30.

Lubar, J. F., & Perachio, A. A. One-way and two-way learning and transfer of an active avoidance in normal and cingulectomized cats. *Journal of Comparative and Physiological Psychology*, 1965, **60**, 46–52.

Lubar, J. F., Perachio, A. A., & Kavanagh, A. J. Deficits in active avoidance behavior following lesions of the lateral and posterolateral gyrus of the cat. *Journal of Comparative and Physiological Psychology,* 1966, **62**, 263–269.

Lubar, J. F., Schastal, C. J., & Perachio, A. A. Non-visual functions of visual cortex in the cat. *Physiology and Behavior,* 1967, **2**, 179–184.

Martin, L. K., Powell, B. J., & Kamano, D. K. Mediation of shock and drug-produced effects on avoidance responding. *Psychonomic Science,* 1967, **9**, 3–4.

McAdam, D. Effects of positional relations between subject, CS, and US on shuttle box avoidance learning in cats. *Journal of Comparative and Physiological Psychology,* 1964, **58**, 302–304.

McAdam, D. W., & Kaelber, W. W. Differential impairment of avoidance learning in cats with ventromedial hypothalamic lesions. *Experimental Neurology,* 1966, **15**, 293–298.

McAllister, W. R., & McAllister, D. E. The inverse relationship between shock intensity and shuttle-box avoidance learning in rats. *Journal of Comparative and Physiological Psychology,* 1971, **74**, 426–433.

McCleary, R. A. Response specificity in the behavioral effects of limbic system lesions in the cat. *Journal of Comparative and Physiological Psychology,* 1961, **54**, 605–613.

McCleary, R. A. Response-modulating of the limbic system: Initiation and suppression. In E. Stellar & J. M. Sprague (Eds.), *Progress in physiological psychology.* Vol. 1. New York: Academic Press, 1966.

Mowrer, O. H. *Learning theory and behavior.* New York: Wiley, 1960.

Moyer, K. E., & Korn, J. H. Effect of UCS intensity on the acquisition and extinction of an avoidance response. *Journal of Experimental Psychology,* 1964, **67**, 352–359.

Moyer, K. E., & Korn, J. H. Effect of UCS intensity on the acquisition and extinction of a one-way avoidance response. *Psychonomic Science,* 1966, **4**, 121–122.

Nadel, L. Dorsal and ventral hippocampal lesions and behavior. *Physiology and Behavior,* 1968, **3**, 891–900.

Olton, D. S., & Isaacson, R. L. Effects of lateral and dorsomedial thalamic lesions on retention of active avoidance tasks. *Journal of Comparative and Physiological Psychology,* 1967, **64**, 256–261.

Olton, D. S., & Isaacson, R. L. Importance of spatial location in active avoidance tasks. *Journal of Comparative and Physiological Psychology,* 1968, **65**, 535–539. (a)

Olton, D. S., & Isaacson, R. L. Hippocampal lesions and avoidance behavior. *Physiology and Behavior,* 1968, **13**, 719–724. (b)

Olton, D. S., & Isaacson, R. L. Fear conditioning, hippocampal lesions, and avoidance behavior. *Communications in Behavioral Biology,* 1969, **3**, 259–262.

Olton, D. S., & Isaacson, R. L. Fear, hippocampal lesions and avoidance behavior. *Communications in behavioral biology,* 1970, **3**, 259–262.

Overmeir, J. B., & Seligman, M. E. P. Effects of inescapable shock upon subsequent escape and avoidance responding. *Journal of Comparative and Physiological Psychology,* 1967, **63**, 28–33.

Papez, J. W. A proposed mechanism of emotion. *Archives of Neurology and Psychiatry,* 1937, **38**, 725–744.

Rich, I., & Thompson, R. Role of the hippocampo-septal system, thalamus and hypothalamus in avoidance conditioning. *Journal of Comparative and Physiological Psychology,* 1965, **59**, 66–72.

Rozin, P., & Kalat, J. W. Specific hungers and poison avoidance as adaptive specializations of learning. *Psychological Review,* 1971, **78**, 459–466.

Saavedra, M. A., Pinto Hamuy, T., & Oberti, C. Auditory and avoidance behavior after extensive and restricted lesion in the rat. *Journal of Comparative and Physiological Psychology,* 1965, **60**, 41–45.

Solomon, R. L., & Wynne, L. C. Traumatic avoidance learning: Acquisition in normal dogs. *Psychological Monographs,* 1953, **67**(4, Whole No. 354).

Theios, J., Lynch, A. D., & Lowe, W. Differential effects of shock intensity on one-way and shuttle avoidance conditioning. *Journal of Experimental Psychology,* 1966, **72**, 294–299.

Thomas, G. J., Hostetter, G., & Barker, D. J. Behavioral functions of the limbic system. In E. Stellar & J. M. Sprague, *Progress in physiological psychology.* Vol. 2. New York: Academic Press, 1968.

Thompson, R. Thalamic structures critical for retention of an avoidance conditioned response in rats. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 261–267.

Thompson, R. L. Effects of lesions in the caudate nuclei and dorsofrontal cortex on conditioned avoidance behavior in cats. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 650–659.

Thompson, R. L., & Mettler, F. A. Permanent learning deficit associated with lesions in the caudate neuclei. *American Journal of Mental Deficiency,* 1963, **67**, 526–535.

Vanderwolf, C. H. Limbic-diencephalic mechanisms of voluntary movement. *Psychological Review,* 1971, **78**, 83–113.

VanHoesen, G. W. MacDougal, J. M., & Mitchell, J. R. Anatomical specificity of septal projections in active and passive avoidance behavior in rats. *Journal of Comparative and Physiological Psychology,* 1969, **68**, 80–89.

Whittleton, J. C., Kostanek, D. J., & Sawrey, J. M. CS directionality and intensity in avoidance learning and extinction. *Psychonomic Science,* 1965, **3**, 415–416.

(Received February 28, 1972)

Psychological Bulletin
1973, Vol. 79, No. 4, 252–259

THE BOGUS PIPELINE:
A NEW *IGNIS FATUUS*?[1]

THOMAS M. OSTROM[2]

Ohio State University

The bogus pipeline has been offered by Jones and Sigall as an improvement on previous rating procedures. Four areas of possible superiority were examined to determine where the pipeline's strengths and weaknesses lie. The pipeline was compared to the conventional rating scale in terms of relative sensitivity, functional comparability, ethical character, and practicality. It was concluded that, with the possible exception of improved sensitivity, the pipeline's superiority was not proven on the basis of the evidence offered.

At night over swampy ground there occasionally appears a phosphorescent light that behaves in a mysterious manner. Sometimes it hovers and sometimes it flits about; and when one approaches it, the light just seems to disappear. The usual explanation of this mischievous light is that it results from spontaneous combustion of swamp gas generated by rotting organic matter. Since this light would occasionally mislead travelers, it came to be known as *ignis fatuus*, a medieval Latin term that translates to "foolish fire."

The history of attitude measurement in social psychology has a similar will-o'-the-wisp character. Just when it seems a satisfactory technique is within our grasp, it manages to elude us. Flaws are invariably found and the search begins anew. In pre-Thurstone days (prior to 1928) many believed that any attempt to quantify attitudes was futile. Yet the need was great and Thurstone, encountering only slight resistance, convinced researchers that attitudes can be measured. But satisfaction with these techniques was short lived, for just as experimental attitude research was getting under way at the end of World War II, doubts began to emerge regarding their fallibility. Early concerns with the possibility of distortion and dissimulation in verbal responses has today become a preoccupation in attitude measurement (Cook & Selltiz, 1964; Rosenthal & Rosnow, 1969). Other techniques, even nonverbal ones such as pupillary response, have had an even more ephemeral existence.

One of the simplest and most commonly employed attitude measurement procedures is the self-rating scale. This procedure has been singled out as the one technique most vulnerable to bias and error (Cook & Selltiz, 1964). A new rating scale procedure has recently been proposed by Jones and Sigall (1971) that, through ingenious development of format and instructions, seemingly offers a solution to some of the more serious response biases. So promising do Jones and Sigall regard this self-rating procedure that they have termed it the "bogus pipeline," a term that suggests the procedure might provide a direct and unobstructed pipeline to the soul.

The basic format of the bogus pipeline is comparable to most self-rating measures. Subjects indicate their attitude on a scale with end labels such as favorable–unfavorable or agree–disagree. Rather than making his rating by pencil, the subject turns a steering wheel that controls a pointer on the rating scale which is mounted on a console in front of him. The crucial difference between this and the typical procedure is the set which a person brings to the rating task. He is told that a machine exists (the electromyograph) that is capable of measuring implicit muscle potentials when attached by electrodes to a person's forearms. When seated with hands on the steering wheel, the subject's first impulse to turn the wheel to a particular rating category could be detected by the machine, even

---

[1] Preparation of this paper was partially supported by Grant GN-534 from the National Science Foundation. The paper benefited from the helpful comments of T. Brock, H. Fromkin, A. Greenwald, A. Gross, J. Keating, B. Latané, and M. Saks.

[2] Requests for reprints should be sent to Thomas M. Ostrom, Social Psychology, 404C West 17th Avenue, Columbus, Ohio 43210.

when the wheel itself is locked in place. By analyzing these implicit muscle potentials, the machine can determine the subject's true attitude on the issue.

Although this cock-and-bull story is apparently greeted with skepticism by most subjects, a set of deceptive demonstrations is presented to convince the subjects that the machine indeed does accurately record their true attitudinal inclinations. This is accomplished by obtaining the subject's self-rating on a variety of topics ahead of time in a manner that arouses no suspicion and then having an accomplice use this information to control the machine's feedback meter when the subject is seated before it.

Once convinced of the machine's infallibility, subjects are told that they are being tested to see how well they understand their own feelings, "whether they were in touch with how they felt." They are asked to estimate their own attitude by using the steering wheel while the feedback meter is concealed from their view. Their self-ratings are to be compared with their true feelings as measured by the machine.

The bogus pipeline is not a test as that term is normally used in attitude and personality measurement. Rather, it is a response format and an instructional set that could be employed in a variety of such tests. It could be used as the agree–disagree scale for Likert attitude tests or for such scales as the Minnesota Multiphasic Personality Inventory (MMPI) and *F* tests. The pipeline is amenable to a forced choice format such as is used in the Rotter I-E scale, or it could be used for the bipolar ratings required by the Semantic Differential.

For the pipeline to be adopted as a standard format for the self-rating of attitudes, its comparability with (or superiority over) existing methods must be established. The data offered to support its relative merit will be discussed in the next several sections.

Relative Sensitivity

A minimum requirement of a new measurement technique is that it correspond to other techniques hypothesized to measure the same thing. Although Jones and Sigall focused on differentiating the pipeline from ordinary rating procedures, it seems of prior interest to establish the concordance of the two approaches.

Convergence of the pipeline with ordinary ratings can be established in two ways. Given that they measure, in part, the same construct, people who score high on one should tend to score high on the other. Second, situational influences that produce a change for one should, if strong enough, lead to change in the other. Convergence across individual differences was shown in a study by Jones, Bell, and Aronson (1972). They correlated attraction toward a confederate as expressed by the pipeline with a 12-item paper-and-pencil rating scale. The within-cell correlations from their two experiments were .66 and .81, a satisfactory level of convergence.

Jones et al. (1972) manipulated whether the confederate had expressed liking or disliking toward the subject, a variable known to reliably affect attraction. This situational variable significantly affected both the pipeline and the rating measures of attraction in both experiments. As would be minimally expected then, these gross situational and individual differences showed a similar pattern for each of the two measurement procedures.

One argument for preferring a newly introduced attitude measure over existing measures is that it contains less error variance in the estimate it provides of each person's attitude. The availability of a more sensitive measure would permit the researcher to detect subtle effects that might otherwise go unrecognized. And, when studying more standard effects, investigators could use smaller sample sizes if the measure were known to be less error ridden.

Several studies were reviewed by Jones and Sigall that indicated the pipeline's sensitivity might be superior to the rating scale. In one paper (Jones et al., 1972), a predicted interaction (between what variables is immaterial for present purposes) was tested in two experiments. Jones and Sigall suggested that the pipeline provided a truer portrayal of the subjects' feelings in these studies because the pattern of means for the pipeline was closer to that predicted than was the pattern of rating measures. But since the predicted interaction did not reach statistical significance for either measure in either experiment, this evidence cannot be given much weight.

Data of a more convincing nature were obtained by Cooper (1971). A dissonance prediction involving a three-way interaction was confirmed when using the pipeline procedure but was not significantly supported when using paper-and-pencil ratings. But as Jones and Sigall noted, there was a possibility of the ratings being influenced by their having followed the pipeline in the sequence of data collection. A preferable test of differential sensitivity would involve half the subjects receiving one dependent measure and half the other measure.

Just such a test was undertaken in an unpublished study by Sigall and Page that was described in Jones and Sigall (1971, p. 358). Subjects evaluated an obnoxious person who was either physically normal or wore leg braces (stigmatized). Half of each group used the pipeline and half used a comparable rating procedure. This stigma manipulation did not produce a significant main effect nor did it interact with the nature of the dependent variable when looking at overall attraction. In terms of detecting differences due to stigma, both measures agreed that no significant differences existed. It appears that neither this study nor the two previously described offer conclusive support for the proposition that the pipeline is more sensitive to attitudinal influences than is the ordinary rating scale.

One potential advantage to the pipeline was shown in this Sigall and Page study. Whereas subjects were reluctant to give an obnoxious person a negative evaluation when using the rating scale, this positivity bias disappeared when the pipeline was used. Although this difference in scale origin may be important for determining the absolute number of pro- and antirespondents, it is irrelevant when testing relational hypotheses.

Functional Comparability

Should the pipeline be shown to possess greater sensitivity than a comparable rating scale, this would not constitute an argument against using the rating scale. It would only mean that larger sample sizes and stronger manipulations would be required with rating procedures than with the pipeline.

But the pipeline may have a more important advantage. To the extent that it offers a purer measure of attitude, the pipeline should provide results more in accordance with reality (and the predictions of accurate theory) than ordinary rating scales. Use of the conventional rating scale in attitude research, then, could lead to spurious validation of erroneous hypotheses. Circumstances may exist where the biases of the rating scale lead to a functional relationship that is different in shape and direction than would be obtained with a less biased measure of attitude. If this attribute of the pipeline could be convincingly demonstrated, it would constitute a powerful argument in the pipeline's favor. Further, such a result would cast serious suspicions on the validity of a great number of past empirical relationships that were based on the rating scale. If it is found that researchers have been misled in their hypothesis testing using conventional rating procedures, then it would become important to devise new measurement procedures, such as the bogus pipeline, that promise a truer view of which independent variables control attitude.

No support for this potential asset was provided in the three already described studies reviewed by Jones and Sigall (1971). Findings that were significant for one measure showed the same pattern of condition means for the other measure. There was no significant tendency for one to show a different functional relationship from the other.

A study by Sigall and Page (1971) tested the prediction that social desirability influences on the rating scale would be eliminated when the pipeline was used. White American subjects rated one of two ethnic groups, Negroes or Americans, on the degree to which each of 22 traits characterized the group. Half of the subjects responded using the pipeline paradigm and half, although using the same steering wheel rating device, were given ordinary rating scale instructions.

Upon first appraisal, some of the findings of this study appeared to support the notion that different functional relations would be obtained, depending on the measure used. For example, one of the traits, honest, was judged more characteristic of Negroes than Americans in the rating condition but less characteristic in the pipeline condition, producing an interaction that was significant. Of the 22 traits judged, 11 were identified by the authors as yielding an interaction supportive of the prediction that the rating scales contain a social desirability bias not shared by the pipeline. These results seemingly verified the ability of the pipeline to eliminate flaws in the rating scale.

Upon closer examination, however, this study does not convincingly substantiate the superiority of the pipeline. Of these 11 interactions, serious consideration can be given to only 6, as the remaining 5 cannot be considered fully reliable ($.20 > p > .05$). The choice of significance level should have been especially conservative in this study as 22 univariate analyses of variance were conducted on nonindependent observations.

Not all of the remaining six interactions were produced by an actual reversal in the direction of the functional relationship between ethnic group and judgment. If a trait is judged more characteristic of one ethnic group than the other, no matter which measure is used, it is wrong to conclude the measures lead to different functional relationships. The two measurement procedures showed comparable relationships for 2 of the traits, leaving only 4 of the original 11 that showed a significant reversal in function direction.

TABLE 1

Traits That Showed an Ethnic Main Effect

Traits	Karlins et al. percentages		Minimum percentage difference
	American	Negro	
Rated most characteristic of Americans			
Aggressive	15	13^a	2
Ambitious	42	13^a	29
Conventional	17	13^a	4
Industrious	23	13^a	10
Intelligent	20	13^a	7
Materialistic	67	13^a	54
Pleasure loving	28	26	2
Practical	12	13^a	—
Progressive	17	13^a	4
Talkative	15^a	14	—
Rated most characteristic of Negroes			
Ignorant	15^a	11	—
Musical	15^a	45	30
Physically dirty	15^a	3	—

[a] The actual percentage is equal to or lower than the recorded figure. Meaningful minimum percentage differences could not be calculated when this percentage was the larger of the two.

A second basis on which the Sigall and Page results should be evaluated is the comparability of their findings with the results of other research on stereotyping. Sigall and Page borrowed their materials from a study conducted 2 years earlier by Karlins, Coffman, and Walters (1969). Karlins et al. examined the percentage of subjects who checked each trait as being characteristic of each ethnic group. The preferred comparison would be between the mean response differences (between Americans and Negroes) obtained by Sigall and Page and the percentage differences for Karlins et al.

Unfortunately, precise comparisons between the two studies are difficult because the Karlins et al. paper did not report the percentages of all traits for both ethnic groups.[3] They only provided data for the adjectives that received the highest percentage of "characteristic" judgments for each ethnic group, and there was little overlap in the top traits listed for Americans and Negroes. Although the actual percentage differences cannot be computed, the *minimum* percentage differences can be derived for 13 of the 22 traits. Nine of these 13 showed a significant ethnic group main effect for Sigall and Page along with either insignificant interactions or interactions that did not involve a reversal of the main effect (see Table 1). For each of these 9 traits, the Karlin's et al. data was in complete accord. So, when both measurement procedures agreed for Sigall and Page, the direction of the relationship exactly coincided with that reported by Karlins et al.

Although Sigall and Page found no significant ethnic group difference or interaction for two of the remaining four comparison traits, an average minimum difference of 11% was observed in the Karlins et al. data. Inspection of the Sigall and Page data shows the pipeline means are in better agreement with the Karlins et al. results than were the rating means for both these traits (see Table 2).

---

[3] These unreported percentages are no longer available from either M. Karlins or T. Coffman (personal communications, December, 1971, and February, 1972, respectively).

Of the remaining two traits for which comparable data were available, one showed a small minimum percentage difference (2%) and the other a somewhat larger difference (11%). Sigall and Page found a significant interaction for each in which the ethnic group differences were in opposite directions for the two dependent measures (see Table 3). For the trait with the smallest minimum difference, sensitive, the direction of its difference coincided with the direction of the rating scale means in the Sigall and Page study. Alternatively, the trait with the larger percentage difference, lazy, was in the same direction as the pipeline means. It appears from these cross-study comparisons that when discrepancies exist between the ethnic group findings of the two studies, the Karlins et al. data are more likely to conform to Sigall and Page's pipeline condition than their rating scale condition.

An index of stereotype favorability was computed by Sigall and Page which showed that the stereotype held of Americans was more positive than it was of Negroes. Although this difference was observed for both the rating and pipeline conditions, it was significantly more pronounced with the pipeline measure. In fact, compared to the pipeline difference of .81 scale units, the rating difference of .04 units was negligible. Karlins et al. also calculated an index of stereotype favorability and they, too, found that the stereotype held of Americans was more positive than was the stereotype of Negroes. So once again a comparison across the two studies shows Karlins et al. data to be more similar to Sigall and Page's pipeline results than to their rating differences.

TABLE 2

TRAITS THAT SHOWED NO ETHNIC MAIN EFFECT OR INTERACTION

Traits	Difference of Negroes over Americans		
	Minimum percentage	Pipeline means	Rating means
Imitative	—	—	—
Happy-go-lucky	12	.40	−.66
Ostentatious	10	.06	−.94
Superstitious	—	—	—
Unreliable	—	—	—

Note. A positive difference indicates the trait was judged more characteristic of Negroes than Americans. No entries were made for traits when meaningful minimum percentage differences could not be calculated.

TABLE 3

TRAITS THAT SHOWED AN INTERACTION IN WHICH ETHNIC DIFFERENCES WERE REVERSED

Traits	Difference of Negroes over Americans		
	Minimum percentage	Pipeline means	Rating means
Honest	—	—	—
Lazy	11	1.40	−.33
Sensitive	2	−.60	1.53
Stupid	—	—	—

Note. A positive difference indicates the trait was judged more characteristic of Negroes than Americans. No entries were made for traits when meaningful minimum percentage differences could not be calculated.

This all suggests a rather startling conclusion. Sigall and Page's pipeline condition did not eliminate bias that existed in previous findings, but rather their rating condition introduced a bias that was not present in the Karlins et al. study. Rather than demonstrating the superiority of the pipeline, they instead have shown that it is possible to present the self-rating task in a way that maximizes biased responding. However, when appropriate precautions are taken with ordinary rating procedures, as was done in the Karlins et al. study, the rating scale and pipeline are comparable.

Usually ratings are made in a setting where no one else observes the judgments being recorded. If the rater is in a large group of respondents, he may well assume that his responses are not going to be singled out for special attention. Often times the subject is not even required to place his name on the rating materials. Finally, except in situations where experimenter bias is being studied, stimulus presentation is usually standardized through using slides, tape recordings, or the printed page.

The rating condition of the Sigall and Page study incorporated none of these precautions. Subjects made their ratings individually in the presence of the experimenter and were required to announce each aloud before proceeding to the next trait. In addition, the experimenter orally presented the 22 traits to each subject, thereby leaving open the possibility of experimenter bias through voice inflection and emphasis.

It must be concluded that the evidence reviewed by Jones and Sigall (1971) does not confirm the notion that the pipeline leads to attitude functions differing in shape or direction from those obtained with more mundane rating procedures.

ETHICAL AND PRACTICAL CONSIDERATIONS

It is conceivable that future research will demonstrate the pipeline's superiority in terms of relative sensitivity and/or the likelihood of its leading to different, and more accurate, functional relations. Even so, this would not mean that the pipeline should routinely be used in attitude research as a replacement for the rating scale. The ethical and practical implications of its use also need to be examined.

One of the debates in fashioning a code of ethics for the conduct of human research centers around the circumstances under which it is permissible to deceive subjects. An apparent point of agreement among the debating parties, however, is that when choosing between two equally appropriate measures, the researcher should always select the one involving the least deception. Although both the pipeline and rating procedures honestly inform the subject that his attitudinal responses are of interest, the ordinary rating scale doesn't employ the deceptive accouterments of the pipeline.

There are two ways the pipeline might distress subjects. If it is used as a response format for questions that are psychodynamically central to a person's well-being, it is conceivable that the pipeline would force him to recognize truths about himself for which such awareness could be harmful. For example, the MMPI contains many items that probe into sensitive areas. For some people, adjustment to life is achieved by denying that certain items apply to oneself. Some extreme attitudes, such as racial prejudice, are viewed as serving important functions for the individual. To the extent that a person's emotional stability is contingent on such beliefs and feelings being suppressed from consciousness, the pipeline, by forcing these responses into consciousness, could disrupt such adaptive patterns.

But for even the mentally healthy subject, the pipeline represents an ethical hazard. It involves an unusual form of double deception. Most deceptions are characterized by the subject merely "going along" with the experimental manipulations. He does not actively disbelieve what he is told, for the scenario is designed to fit in with the subject's prior conceptions of the world. On the other hand, the pipeline presents the subject with an unbelievable situation, one that goes against his common sense. Yet by ingenious duplicity, the experimenter is able to overcome the subject's initial skepticism.

Then comes the debriefing. With many deceptions, a subject's self-esteem is little affected since the situation seemed reasonable and plausible from the very beginning. Anyone would have taken it the way he did. But it is conceivable that the subject's reaction to discovering the deception of the pipeline is quite different. In accepting the deception, he must acknowledge that his own ability to defend his prior notions of reality was lacking and that the experimenter was able to put one over on him. The subject might prefer to believe at this point that the machine really could measure his attitudes rather than his having to yield once again to a new set of arguments presented him by the experimenter. To go along with those arguments would mean that he had twice been unable to defend convictions that he honestly and firmly held. Jones and Sigall (1971) reported that subjects were "quite surprised" to learn that the pipeline was bogus. Subjects may leave the experiment not knowing which of the two stories to believe.

Whether this deception actually does affect subjects' self-esteem or faith in human nature (or faith in the psychologists' nature) should be empirically examined. Indeed, it may well be possible to develop debriefing procedures that not only nullify negative reactions to the pipeline, but convert them to positive growth experiences.

Consider next, the matter of the pipeline's practicality. Unlike the rating scale, the pipeline presents difficulties in constructing the apparatus, transporting the apparatus, running more than one subject at a time, and requiring excessive time for instructing each subject on how to use the apparatus and how to believe in it. But these are tolerable liabilities; experimenters should be willing to put in the extra effort if the pipeline's superiority can be established.

There is one practical problem that imposes a potentially serious limitation on the pipeline. If the pipeline becomes routinely adopted, the nature of its deception would probably become widely known thereby making it difficult to obtain truly naive subjects. This would be especially true for the undergraduate culture, a relatively closed system that can be quite efficient in transmitting information of this kind to future generations. This requirement of subject naiveté, vital to the pipeline's success, is not shared by ordinary rating procedures.

This feature of the pipeline presents a nettling problem to its advocates. If it is true that widespread adoption of the paradigm, such as is facilitated by its publication in the *Psychological Bulletin*, will lead to awareness of its deceptive nature by the undergraduate subject pool, then the pipeline will become unusable for such subjects. And the more often it is used, the more rapidly this point of invalidity will be reached. There is no way of concealing its deceptive features short of withholding this information during debriefing. Even then, knowledge of the technique may be disseminated in other ways. Articles in the *New York Times Magazine*, *Psychology Today*, or the behavior section of *Time* magazine could hobble the technique for years to come.

CONCLUSIONS

Whenever a new measurement procedure is proposed, it is necessary to examine fully its strengths and weaknesses. The task of test evaluation is made somewhat easier when a widely used technique, such as the rating scale, exists as a standard of comparison. Instead of having to make an absolute judgment of good or bad, one can make the comparative decision of better or worse.

Four bases of comparison have been outlined: relative sensitivity, functional comparability, ethical character, and practicality. Evaluating the pipeline's merit in comparison with conventional rating formats on these grounds leads to the conclusion that the pipeline's superiority has not yet been proven.

The data reviewed by Jones and Sigall are encouraging, but not conclusive regarding the relative sensitivity of the pipeline. There was some suggestive evidence that it may contain more true variance than the rating scale and that its scale origin might be more veridically located.

Whereas the pipeline may prove to be superior on sensitivity grounds, it is decidedly less preferable than the rating scale on both ethical and practical grounds. If the extra sensitivity is not needed for testing the hypotheses under consideration, it would appear preferable to continue using a traditional rating format.

Clearly the most important potential contribution of the pipeline lay in demonstrating that it leads to different functional relationships than the rating scale. If the pipeline could provide tests of hypotheses unobtainable with other measures, then researchers should be more willing to cope with its ethical and practical drawbacks.

It would seem, then, that establishing functional noncomparability would be of primary concern, especially in view of the possibility that the pipeline has a limited half-life due to people becoming aware of its deceptive nature. If such noncomparability can be established, then additional research is warranted to determine (*a*) whether the pipeline contains different or fewer response biases than the rating scale and/or (*b*) whether it measures a different property of attitude than is assessed by the rating scale.

If the pipeline is shown to be different from the rating scale, an alternative to the pipeline should be sought that, while possessing the pipeline's essential properties, would not have the ethical and practical deficiencies of the pipeline inherent in it. If the primary differences were found to be in regard to response biases, then other instructional sets that encourage honesty in a way comparable to the pipeline could be tested. On the other hand, if the differences between the two measures were due to their tapping different facets of attitude, then instructional sets could be prepared that sensitize the subject to the specific aspect of attitude that the investigator wished to elicit. Successful attempts at such differential operationalization of affective, conative, and cognitive components of attitude have been achieved by Kothandapani (1971) and Ostrom (1969) when employing rating techniques.

Jones and Sigall have developed a very interesting and potentially informative methodological device. One approach they have employed for establishing its value, conducting parallel experiments that differ only in the dependent measures being compared, serves as a model research design for examining dependent variable differences. It is certainly superior to most correlational approaches that have been proposed (e.g., Gulliksen, 1968). Few other critics of self-rating scales in attitude measurement have initiated such a careful approach to examine the empirical implications of their criticisms. One hopes that Jones and Sigall, as well as other interested investigators will continue this program of research on the pipeline's differences from the rating scale. In such research, more conventional formats of the rating scale should be used rather than the version adopted by Sigall and Page (1971). The pipeline should be compared with the best, rather than the worst, of the conventional rating scale formats.

If pipeline research focuses on the two problem areas outlined above, social psychology could yield substantial benefit in terms of understanding the components of attitude and/or response biases in rating measures. This appears to be a more profitable use of the pipeline's possibly limited half-life than employing it in situations (such as testing dissonance predictions) where an ordinary rating scale should be adequate. One important feature of such research is that it culminates in the obsolescence of the very measure it seeks to validate. This temporary and restricted use of a research paradigm could be profitably applied to other questionable methodologies.

REFERENCES

Cook, S. W., & Selltiz, C. A multiple-indicator approach to attitude measurement. *Psychological Bulletin,* 1964, **62**, 36–55.

Cooper, J. Personal responsibility and dissonance: The role of foreseen consequences. *Journal of Personality and Social Psychology,* 1971, **18**, 354–363.

Gulliksen, H. Methods for determining equivalence of measures. *Psychological Bulletin,* 1968, **70**, 534–544.

Jones, E. E., Bell, L., & Aronson, E. The reciprocation of attraction from similar and dissimilar others. In C. McClintock (Ed.), *Experimental social psychology.* New York: Holt, 1972.

Jones, E. E., & Sigall, H. The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin,* 1971, **76**, 349–364.

Karlins, M., Coffman, T. L., & Walters, G. On the fading of social stereotypes: Studies in three generations of college students. *Journal of Personality and Social Psychology,* 1969, **13**, 1–16.

Kothandapani, V. Validation of feeling, belief, and intention to act as three components of attitude and their contribution to prediction of contraceptive behavior. *Journal of Personality and Social Psychology,* 1971, **19**, 321–333.

Ostrom, T. M. The relationship between the affective, behavioral, and cognitive components of attitude. *Journal of Experimental Social Psychology,* 1969, **5**, 12–30.

Rosenthal, R., & Rosnow, R. L. *Artifact in behavioral research.* New York: Academic Press, 1969.

Sigall, H., & Page, R. Current stereotypes: A little fading, a little faking. *Journal of Personality and Social Psychology,* 1971, **18**, 247–255.

(Received March 27, 1972)

Psychological Bulletin
1973, Vol. 79, No. 4, 260–262

WHERE THERE IS *IGNIS*, THERE MAY BE FIRE

EDWARD E. JONES[1]
Duke University

HAROLD SIGALL
University of Rochester

The "bogus pipeline" is defended, not as an all-purpose substitute for conventional rating scales, but as a procedure useful in certain settings to inhibit social desirability influences or to explore the affective components of attitudes. Previous studies do show reliable differences between bogus pipeline estimates and conventional ratings, though the interpretation of these differences remains equivocal. Issues of ethics and practicality are discussed with the conclusion that difficulties with the bogus pipeline are far from unique to this procedure.

We confess a decided ambivalence toward Ostrom's (1973) critique of our "bogus pipeline" paper. We do not know whether to be pleased with the conclusion that "Jones and Sigall have developed a very interesting and potentially informative methodological device [p. 259]" or to be dismayed by being referred to as "rotting organic matter" generating "swamp gas."

The reader who has not seen the original paper might be persuaded by Ostrom that we have urged the abandonment of conventional rating scales in favor of bogus pipeline procedures. Thus, we note Ostrom's reference to the issue of adopting the pipeline "as a standard format for the self-rating of attitudes [p. 253]." And, again, there is a reference to whether or not the "pipeline should routinely be used in attitude research as a replacement for the rating scale [p. 257]." The straw man introduced by such remarks is inconsistent with the thrust of our paper. At no point did we advocate, or even consider, the routine substitution of the pipeline for conventional rating scales. In one of our concluding remarks (Jones & Sigall, 1971), we emphasized that "it is important not to embrace too hastily the proposition that the bogus pipeline is necessarily a better or truer measure of attitude than the standard devices [p. 363]." It was certainly our intention to be extremely cautious in claiming advantages for the bogus pipeline, and our effort was to indicate by example some of the particular circumstances where it might produce different results. We completely agree with Ostrom that the differences between rating scale and pipeline results cannot at this stage be rigorously interpreted.

We do not agree, however, that no important differences have been demonstrated. Ostrom draws a sharp distinction between "sensitivity" and "functional comparability." We believe that such a distinction is very difficult to maintain, or at least that its importance has been exaggerated. Significant crossover interactions are, of course, dramatic. In our suggestion that the pipeline, in some conditions, will convey a more accurate picture of attitudes, we make no claim that pipeline results always or often *reverse* rating scale results. A number of studies show significant effects with the pipeline measure and not with conventional rating scales (e.g., Cooper, 1971). To argue that such differences can be overcome by simply increasing the number of subjects in the rating condition is fallacious. If the rating results of two experimental conditions are practically identical, it is hard to see how the addition of further subjects will automatically guarantee a significant difference. Also, a significant crossover interaction *was* obtained in a study by Jones and Wein (1972) that compared bogus pipeline estimates with ratings in judging a person who expressed increasing or decreasing opinion similarity to the subject. In general, we do not see why crossover interactions are any more necessary for testing relational hypotheses than other interactions and main effects.

Ostrom spends considerable space comparing the Sigall and Page (1971) study with an earlier study by Karlins, Coffman, and Walters (1969). His main argument here is that the latter results fall in the same pattern as the bogus pipeline results in Sigall and Page whereas the conventional rating results do not. Although cross-study comparison ignores the "vast difference in approach and procedure [Sigall & Page, 1971, p. 25]," we agree that the rating scale results may have been contaminated by social desirability standards. Indeed, Sigall and Page (1971) explicitly acknowledge this possibility and emphasize that "the evidence we have gathered, which suggested that social desirability needs affected our rating scale subjects, does not in any way prove that such demand characteristics were operative in any of the Princeton studies [p. 253]." What the study does show is that the bogus pipeline inhibits the influence of social desirability norms and produces stereotype scores that are higher for Americans and lower for Negroes.

---

[1] Requests for reprints should be sent to Edward E. Jones, Department of Psychology, Duke University, Durham, North Carolina.

Ethics and Practicality Issues

We certainly agree that the use of the bogus pipeline raises complex ethical issues and endorses Ostrom's conclusion that the pipeline should not routinely replace the conventional rating scale. We also accept the general context within which Ostrom chooses to discuss research ethics, namely, that the gain in useful scientific information about human nature should outweigh the potential dangers involved for the individual subject.[2]

As we see it, the basic danger involved for the individual subject is the distress conceivably caused by an invasion of his privacy. Similar invasions are also implicit in the administration of projective tests and other personality probes whose interpretation is not self-evident. If our assumptions about the efficacy of the bogus pipeline are correct, subjects will be induced to reveal attitudes or attitudinal components that they would prefer to conceal. Certainly in most cases, and with most attitudes, this is less likely to cause distress than many experimental situations in which the subject is induced to prevaricate, to cheat, or to bring harm to another subject. The bogus pipeline subject is *not* forced to acknowledge to others or to himself that he would respond differently to a conventional rating scale. Even if the subject becomes aware of this possibility, Ostrom himself admits that a sensitive debriefing can result in a "positive growth experience." In our own work we have run across very few subjects who were even mildly upset by the bogus pipeline procedure.

Ostrom argues that the bogus pipeline involves an unusual form of "double deception," but nothing he subsequently says justifies the use of such a label. Somehow, unlike other deception experiments, the bogus pipeline allegedly convinces people against their will. Distress is caused by the subject's realization that he has been "taken in" by the experimenter. We fail to see how a bogus pipeline experiment involves more, or a different kind of, deception than any other experiments attempting to convince subjects that what is false is really true. We are particularly surprised by the suggestion that subjects might prefer to believe that the machine really works and will not, therefore, accept the debriefing. On the contrary, subjects can be easily led to understand the deception and the reasons for it; they are normally relieved to learn that such a measuring device is still beyond the grasp of psychological science. Still, the ethical issues associated with use of the bogus pipeline cannot be summarily dismissed. Each investigator will have to ask himself (and hopefully his colleagues) whether the probable knowledge to be gained is worth the potential of distress to subjects.

The practicality issue is only an issue if one assumes that every social psychologist will rush to build his own pipeline and that bogus pipeline studies will sweep the campuses of America. While such a fantasy flatters the technique, we have urged that the bogus pipeline should only be used when important theoretical issues are involved, when subjects are otherwise likely to resist disclosure of their attitudes, or when the experimenter is interested in tapping different attitudinal components. Overconcern with the practicality issue would long ago have ruled out conformity studies, self-esteem manipulations, obedience research, in short, any experimental program that involves fairly standardized deception techniques. It is hard to argue that our knowledge of human nature has not been advanced by such deception research.

---

[2] Such a relativistic context is a guiding assumption in the latest draft prepared by the American Psychological Association's Ad Hoc Committee on Ethical Standards for Psychological Research.

In conclusion, we agree with Ostrom that the meaning of bogus pipeline results must be clarified by further research. We disagree that no important differences between rating scales and pipeline estimates have been revealed to this point. We do not see inherent in the bogus pipeline paradigm any ethical or practical difficulties that would single out the paradigm and distinguish it in this respect from other deception research. And finally, we believe that Ostrom's message is a confusing one. On the one hand, he argues that nothing has turned up to justify a claim for the superiority of the bogus pipeline so that the ethical prohibition should stand. On the other hand, he exhorts us (and others) to continue to probe through the use of further research for any differences that might exist. It seems to us that there is an inherent contradiction in these exhortations.

REFERENCES

COOPER, J. Personal responsibility and dissonance: The role of foreseen consequences. *Journal of Personality and Social Psychology*, 1971, **18**, 354–363.

JONES, E. E., & SIGALL, H. The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, 1971, **76**, 349–364.

JONES, E. E., & WEIN, G. Attitude similarity, expectancy violation, and attraction. *Journal of Experimental Social Psychology*, 1972, **8**, 222–235.

KARLINS, M., COFFMAN, T. L., & WALTERS, G. On the fading of social stereotypes: Studies in three generations of college students. *Journal of Personality and Social Psychology*, 1969, **13**, 1–16.

OSTROM, T. M. The bogus pipeline: A new *ignis fatuus? Psychological Bulletin*, 1973, **79**, 252–259.

SIGALL, H., & PAGE, R. Current stereotypes: A little fading, a little faking. *Journal of Personality and Social Psychology*, 1971, **18**, 247–255.

(Received March 27, 1972)

Psychological Bulletin
1973, Vol. 79, No. 4, 263–270

SOCIOECONOMIC STATUS AND TRADITIONAL TREATMENT APPROACHES RECONSIDERED[1]

RAYMOND P. LORION[2]

University of Rochester

In view of scarce mental health services and previous research with traditional therapies, innovative treatment approaches are employed increasingly with low-income patients. Pertinent data must be carefully considered lest the present trend produce separate and unequal services for various social groups. While socioeconomic status appears to be a significant correlate of acceptance for, and duration of, individual psychotherapy, it does not relate to treatment outcome. Clarification of this discrepancy requires a more precise identification of the social groups comprising the "lower-classes"; a clearer understanding of the environmental demands faced by such patients; and an objective assessment of specific socioeconomic status factors that may, or may not, exclude a given patient from a given treatment. Resolution of such issues is necessary for a realistic allocation of available mental health services.

A warning sounded by Hollingshead and Redlich (1958) more than a decade ago remains equally pertinent today:

> The suggestion that different social classes receive different treatment for mental illness may come as a shock, but to repress facts because they are distasteful and incongruent with cherished values may lead to consequences even more serious than those we are trying to escape by substituting fantasy for reality [p. 4].

Given the critical discrepancy between the demand for mental health services (Zax & Cowen, 1972) and their current or projected availability (Albee, 1969), it is becoming increasingly important to identify patient populations for whom given therapeutic interventions are effective. Accordingly, Strupp and Bergin (1969) identify as the principal empirical question for future research in therapy: "What specific therapeutic interventions produce specific changes in specific patients under specific conditions [p. 20]?" While this is an extremely complex question, it must be resolved if services are to be allocated realistically.

To plan optimal mental health services for low-income individuals, research findings dealing with their experiences in applying for outpatient treatment at mental health facilities must be critically assessed. Notwithstanding the availability of relevant data, programs continue to be designed and operated, based on strongly held views without empirical grounding. Given the recent proliferation of innovative treatment approaches for the poor (e.g., Guerney, 1969; Minuchin, Montavo, Guerney, Rosman, & Schumer, 1967; Small, 1971; Sobey, 1970), therapists' treatment options have increased dramatically. Economic factors decrease the likelihood that such patients will be assigned to classical therapy approaches. Lerner (1972) raises several important issues relative to such decisions and emphasizes the need for careful considerations of therapeutic effectiveness lest two separate yet unequal treatment systems evolve —one for the poor and another for the well-to-do. In this context, the present paper reviews findings dealing with low-income patients applying for mental health services at clinics emphasizing traditional treatment approaches.

CLASS IV VERSUS CLASS V

Defining a group's characteristics is an essential precondition to considering its experiences! The most widely used instrument for determining socioeconomic status in psychotherapy research has been the Hollingshead Index of Social Position, developed for use in the New Haven community study (Hollingshead & Redlich, 1958). Initially, the

---

[1] The author acknowledges the invaluable critical comments of Emory Cowen and Lonnie E. Mitchell throughout the preparation of this manuscript.

[2] Requests for reprints should be sent to Raymond P. Lorion, Department of Psychology, University of Rochester, Rochester, New York 14627.

TABLE 1

OCCUPATIONAL, EDUCATIONAL, AND FAMILIAL CHARACTERISTICS OF FIVE SOCIOECONOMIC STATUS LEVELS

Class	Occupational level	Educational level	Family structure
I	Salaried positions in policy-making executive level; private-practice professionals	Professional degrees; A.B. level and beyond	Modal nuclear family of parents and children, with stability encouraged
II	Salaried positions in business and professions; minor professionals included	A.B. level or partial college	Modal nuclear family of parents and children with stability encouraged
III	"Middle-class" administrative, clerical, sales, technical, and semiprofessional positions	High school diploma	Modal nuclear family of parents and children with stability encouraged
IV	"Working-class" skilled and semiskilled manual occupations in unionized trades and industries	High school or technical school diploma with some below tenth grade	Modal nuclear family often three generations, instability more common than I–III
V	"Poor" semiskilled and unskilled manual occupations nonunionized with irregular employment	High school diploma infrequent with many not completing eighth grade	Modal nuclear family extended to 3–4 generations; divorce, separation, and instability common

Note. The socioeconomic status levels are referenced in Hollingshead and Redlich, 1958.

Index defined class status in terms of three variables: area of residence, occupation, and educational background. Later, it was revised to exclude residence (Hollingshead, 1957) and a correlation of .97 between these two forms has been reported (Myers & Bean, 1968). The Index identifies five socioeconomic status levels, the highest being Class I; the lowest, Class V. Table 1 summarizes the major characteristics of the classes. Although the Index was developed specifically for the New Haven community in the 1950s, it is assumed to have some continuing generality. Future revision of this and related indices, however, should reflect the profound educational, economic, and social changes of the last two decades—changes felt mostly by low-income groups. Recent studies of low-socioeconomic-status groups continue to rely on the Index as the critical socioeconomic status measure, without correction for ongoing social change. Findings to be reviewed must be evaluated with this limitation in mind.

The two low-income groups (Classes IV and V) have generally been perceived as similar on the basis of certain commonly shared characteristics. Both are typically involved in manual occupations, are removed from the middle- and upper-class "prestige race," are "down-to-earth," are pragmatically oriented (Miller & Riessman, 1969), tend to live in informal, comfortable relationships with others, and appreciate the affectionate bite of humor (Gans, 1962). These communalities notwithstanding, it is misleading to ignore significant differences between these two groups and thus to consider them as a homogeneous segment of the population.

Miller and Riessman (1969) suggest that the unskilled, irregular, Class V worker lacks the disciplined, structured, working-class approach. He is less able to cope with life's demands because of its seeming hopelessness. Although higher socioeconomic status groups neither admire nor understand the Class IV working man, they can appreciate the products of his skill and efforts. The poor, however, are objects of the prejudice and lack of concern of all other classes, including Class IV. Class V members are far less likely to attain middle-class goals and possessions than are working-class members. Labor unions, a prime means of organization and political power for the working class, are often inaccessible to the Class V worker. The combination of economic insecurity and familial instability limits the poor to a crisis reactive existence (Miller, 1964) not generally experienced by Class IV members. Class V existence is further characterized by a lack of control over the forces governing an individual's existence (Alinsky, 1967; Haggstrom, 1964).

Differentiations among Class V members are themselves extensive (Riessman, Cohen, & Pearl, 1964). The poor include unskilled and irregularly employed whites and nonwhites, migrant farm workers, numerous minority group members, ghetto dwellers, and tenant farmers. Sarbin (1970) notes that Class V consists of both the respectable and the degraded poor, the latter being continual participants in a culture of poverty. Inclusion in such a culture identifies one as "disinherited" and thereby as "not included in the collectivity that makes up the 'real' society of 'real' people [Rainwater, 1970, p. 9]."

While the preceding account is descriptive and anecdotal, it suggests the need to avoid generalized references to the "lower classes." Classes IV and V include different groups with differing life styles—a distinction typically ignored in the findings reviewed.

Fishman (1969) and Thomas and Sillen (1972) are strongly critical of the relative absence of empirical data about the feelings, motivations, and behavior of the poor, both white and nonwhite. A similar lack is also true for the working class. While Gans (1962) and Minuchin et al. (1967) provide some insights into the life-styles of the poor, more rigorous, integrated evaluations of ego strengths, problem-solving capacities, and cognitive organizations is needed for these groups. Allen (1970) identifies needed areas of research with the poor and emphasizes the contribution that such studies would make. At present, there is little scientific data about such matters.

While many innovative treatment approaches are ahistorical (Small, 1971; Sobey, 1970; Yamamoto & Goin, 1965), traditional therapies tend more to consider knowledge of the cultural background of patients in trying to understand the evolution of a patient's disorder (Savitz, 1952). Lerner (1972) questions whether in the absence of such knowledge it is possible for the therapist to empathize optimally with a patient and hence establish therapeutic relationships comparable to those observed with middle- and upper-class patients. This caution must be seriously considered in an assessment of previous research findings with low-income groups. Furthermore, therapeutic contracts and goals must recognize and operate within the realistic limits of the patient's environment (Chess, Clark, & Thomas, 1953; Green, 1946). To ignore the social reality of a patient's existence while attempting to implement change in therapy may be self-defeating. Failure may often follow when the environment into which a patient must return will not accept such change (Thomas & Sillen, 1972); at such time it may be a sign of health for a patient to refuse change and to withdraw from treatment.

Schneiderman (1965) suggests that it might be helpful to the low-income patient if the professional assesses his adaptive and problem-solving capacities in relation to the situations in which they occur. Without such knowledge, the professional may be subject to ethnocentric prejudgments about the patient's behavior. Healthy middle-class solutions may not be healthy for the problems of low-income people. Cohen (1964) and Fishman (1969) have both suggested that some forms of behavior, deviant by middle-class standards, are the *only* available mastery outlets for low-income individuals and *must* be used to avoid loss of self-esteem or a sense of degradation. Grier and Cobbs (1968) perceive few other alternatives as typically available to ghetto dwellers. For the poor, aggression and acting out against authority figures are important coping devices in dealing with certain ego threats (Grey, 1969). Such behaviors may be necessary substitutes for the low-income person's impulses to cry, plead, or talk things over—each socially unacceptable (Fishman & McCormack, 1969; Silberman, 1964).

SOCIOECONOMIC STATUS AND TREATMENT

Although Classes IV and V have been described as the most psychiatrically "impaired" segment of society (Srole, Langer, Michael, Opler, & Rennie, 1962), they have the least chance of obtaining professional help in outpatient facilities (Hollingshead & Redlich, 1958; Srole et al., 1962). Kadushin (1969) reports that less than 5% of all patients seen at Manhattan Mental Health Clinics were members of minority groups—a proportion far below general population rates for that area. In the studies reviewed below, socioeconomic status correlates significantly and negatively with acceptance for and duration of individual psychotherapy, with experience level of assigned therapist, but not with a patient's diagnostic category or source of referral. These findings take on greater significance in that the data were drawn from records of clinics in which ability to pay was *not* a condition for treatment.

A number of reports from the New Haven studies indicate that low socioeconomic status is negatively related to acceptance for individual psychotherapy (Hollingshead & Redlich, 1958; Redlich, Hollingshead, Roberts, Robinson, Friedman, & Myers, 1953; Robinson, Redlich, & Myers, 1954). Similarly, the case records reviewed by Schaffer and Myers (1954), Myers and Auld (1955), Redlich (1958), and Myers and Schaffer (1967) confirm the importance of socioeconomic status in treatment decisions. At the New York Veterans Administration Regional Office, being assigned to psychotherapy, rather than to a somatic treatment, was significantly correlated with a patient's socioeconomic status level (Bailey, Warshaw, & Eichler, 1960). Therapist's ratings of patient motivation and acceptability for treatment related significantly and negatively to socioeconomic status for the 22 Veterans Administration clinics surveyed (Bookbinder & Gusman, 1964; Raskin, 1961). Rosenthal and Frank (1958) found through an analysis of case records of 384 applicants for outpatient treatment that psychiatrists refer to therapy persons most like themselves, that is, whites rather than nonwhites and those in the upper rather than in the lower income range. Class V applicants were least likely to be accepted for psychotherapy at the University of Utah Medical School Clinic (Codes, Branch, & Allison, 1962). Shader (1970) noted that low-socioeconomic-status applicants to a walk-in clinic were significantly more likely to receive drugs than individual psychotherapy.

Attrition rates among low-income patients assigned to individual psychotherapy were significantly higher than for their middle- and upper-class controls. Myers and Roberts (1959) observed that low-income applicants accepted for therapy either quit in disgust or were dropped as failures. Low-socioeconomic-status patients seen in Veterans Administration settings terminated prematurely significantly more often than did patients from other socioeconomic status levels (Dengrove & Kutash, 1950; Sullivan, Miller, & Smelser, 1958; Winder & Hersho, 1955).

At the Phipps Psychiatric Clinic, social class and attrition were significantly and negatively related (Imber, Nash, & Stone, 1955; Frank, Gliedman, Imber, Nash, & Stone, 1957; Rosenthal & Frank, 1958; Imber, Frank, Gliedman, Nash, & Stone, 1956). Neither residents nor patients could be "pressured" by clinic administrators to maintain a relationship that was mutually unsatisfying. Nash, Hoehn-Saric, Battle, Stone, Imber, and Frank (1965) report that socioeconomic status related positively to therapist's ratings of patient attractiveness, ease of establishing rapport, and good prognosis, each of which related significantly to treatment duration. A review of 300 case records further confirms the negative relation between socioeconomic status and attrition rates (Katz, Lorr, & Rubenstein, 1958; Lorr, Katz, & Rubenstein, 1958). In addition, Schneiderman (1965), working in a Family Service Clinic, found that with declining socioeconomic status level, there was a greater liklihood that the social worker, rather than the patient, would initiate termination.

Social class has thus been shown to be an important determinant of acceptance for, and duration of, individual psychotherapy (Sanua, 1966). Low-socioeconomic-status applicants are significantly less likely to be assigned to individual treatment, and if assigned, are more likely to terminate prematurely. These relations have been observed across a variety of settings in different parts of the country. Generally speaking, low-socioeconomic-status contraindicates individual therapy.

The prepotency of socioeconomic status in treatment decisions is reflected in reports that assignment to therapy is more a function of social class than of diagnosis (Myers & Auld, 1955; Myers & Schaffer, 1967; Robinson, Redlich, & Myers, 1954). The absence of a relation between treatment disposition and diagnosis has been observed in the records of Veterans Administration clinics (Bailey, Warshaw, & Eichler, 1960), of the Phipps Clinic (Imber et al., 1955, 1956), of the New York State Psychiatric Institute (Budner, Escover, & Malitz, 1964), and of the University of Chicago Medical School (Heine & Trosman, 1960). Among diagnostically comparable low-socioeconomic-status applicants for therapy at the Los Angeles General Hospital, unemployment (Yamamoto & Goin, 1965) or minority group membership (Yamamoto, James, Bloombaum, & Hatter, 1967) caused them to be seen as even less appropriate for therapy.

There has been no systematic verification of Schaffer and Myers' (1954) finding that low-income patients are disproportionately assigned to inexperienced therapists. Studies of low-socioeconomic-status patients in individual therapy have almost exclusively used residents, medical students, or psychology interns as therapists; senior staff members apparently did not treat Class IV or V patients in these studies. This datum indirectly supports the view that socioeconomic status is a determinant of who will treat a patient. While it has not yet been established that experienced therapists produce more positive outcomes (Bergin & Garfield, 1971; Lerner, 1972; Meltzoff & Kornreich, 1970) Sanua's (1966) argument that lower income patients may not appreciate inexperienced therapists, deserves consideration.

The tendency to see Class IV and V individuals as part of a blurred, homogeneous, "lower-class" obscures essential empirical questions. It is valid to ask whether treatment outcomes are the same for these two groups. All too often comparative analysis between Classes IV and V has been impossible because samples studied were not even coded with this distinction in mind. That Class IV applicants differed significantly from Class V applicants in acceptance rates for therapy (Coles, Branch, & Allison, 1962) suggests that Class IV and V treatment experiences are not identical. Furthermore, unemployment and minority group membership, characteristics of Class V, significantly differentiated low-income applicants accepted for therapy (Yamamoto & Goin, 1965; Yamamoto et al., 1967). Although far from conclusive, such data underscore the need to identify more systematically, specific socioeconomic-status-related factors involved in treatment decisions.

Conclusions about entering or remaining in therapy drawn from the studies reported, do not establish a relation between socioeconomic status and treatment outcome (Truax & Carkhuff, 1967). Katz, Lorr, and Rubenstein (1958) did not find social class differences among patients designated as "successes" by their therapists; all socioeconomic status levels were represented approximately equally among successful patients. Similar observations have been reported by Rosenthal and Frank (1958), Brill and Storrow (1960), and Coles, Branch, and Allison (1962). Albronda, Dean, and Starkweather (1964) provide suggestive evidence supporting Frank's (1961) view that overcoming reality factors associated with continuation in therapy for low-socioeconomic-status patients may reflect an extremely strong motivation for change. After the eleventh interview, low-income patients responded slightly better to treatment than did higher socioeconomic status controls. The potential for positive outcome in traditional treatment approaches with low-income patients has been documented in an important study reported by Lerner (1972). In addition, the economic feasibility of providing such treatment was also supported by Lerner's (1972) findings.

The studies reviewed also ignore the fate of those rejected for, or rejecting, treatment. Brandt (1964) and Riess and Brandt (1965) provide data suggesting that significant numbers of "rejectees" enter treatment elsewhere. Does the low-income "rejectee" apply and meet a similar fate at another mental health facility? Is he accepted for treatment elsewhere? Conceivably, the initially rejected low-income applicant may become part of a pool of repeatedly rejected applicants, accepted low-socioeconomic-status applicants, or successful therapy cases. Data bearing on such questions do not exist!

The observed discrepancy between acceptance and outcome data for low-income therapy patients has generated a number of explanations—few empirically tested. Truax and Carkhuff (1967) suggest that the acceptance data typically reflect prejudices common to the field. Gundlach and Geller (1958), Miller and Mishler (1959), and Koegler and Brill (1967) conclude that selection factors related to socioeconomic status more often meet therapist needs than treatment requirements. Fierman (1965) dismisses all socioeconomic status limitations to treatment as myths with no basis whatsoever in reality. In his view, the very factors used to rationalize exclusion of low-income individuals from treatment justify their need for therapy.

In a satirical essay, Adams and McDonald (1968) point up an analogy between the reception of low-income applicants at clinics and the "cooling-out" practices that con men use on their victims. They suggest that the "real purpose" of the intake interview is to amass evidence that the poor are neither treatable nor motivated. Similarly, McMahon (1964) argues that the fact that therapists "permit" so many of the poor to refuse treatment, without allowing them to understand what they are rejecting, is tantamount to a "*de facto* practice of disengagement from the poor [p. 287]." Chessick (1971) states that treatment decisions based merely on a patient's economic qualifications violate a basic human right "to be able to receive the best possible medical care regardless of socioeconomic status [p. 37]."

Although the current social climate favors the practice of indicting policymakers for apparent prejudicial distribution of services (Thomas & Sillen, 1972), the soundness of explaining the findings on this basis alone can be challenged. In fact, the data reviewed indicate that Class IV and V applicants simply do not receive, or continue in, individual psychotherapy. The defensibility of the view that rejected or prematurely terminating patients would have profited from therapy as much as "successful" low-socioeconomic-status patients do, is questionable. Furthermore, there is no evidence indicating that terminators did not, in fact, profit to some degree. There are far too many individual difference variables to justify such conclusions without solid objective data. Clearly, it is inequitable for socioeconomic status factors alone to determine who should receive treatment and in what form. The discrepancy between acceptance and outcome data underscores this fact. Equally, it is unreasonable to assume that alternative forms of treatment do disservice to those not receiving individual therapy.

REFERENCES

ADAMS, P. L., & MCDONALD, N. F. Clinical cooling out of poor people. *American Journal of Orthopsychiatry,* 1968, **38**, 457–463.

ALBEE, G. W. The relation of conceptual models of disturbed behavior to institutional and manpower requirements. In F. N. Arnhoff, E. A. Rubenstein, & J. C. Speisman (Eds.), *Manpower for mental health.* Chicago: Aldine, 1969.

ALBRONDA, H. F., DEAN, R. L., & STARKWEATHER, J. A. Social class and psychotherapy. *Archives of General Psychiatry,* 1964, **10**, 276–283.

ALINSKY, S. D. The poor and the powerful. *Psychiatric Research Reports,* 1967, **21**, 22–28.

ALLEN, V. L. (Ed.) *Psychological factors in poverty.* Chicago: Markham, 1970.

BAILEY, M. A., WARSHAW, L., & EICHLER, R. M. Patients screened and criteria used for selecting psychotherapy cases in a mental hygiene clinic. *Journal of Nervous and Mental Disease,* 1960, **130**, 72–77.

BERGIN, A. E., & GARFIELD, S. L. *Handbook of psychotherapy and behavior change.* New York: Wiley, 1971.

BOOKBINDER, L. J., & GUSSMAN, L. J. Social attainment, premorbid adjustment, and participation in inpatient psychiatric treatment. *Journal of Clinical Psychology,* 1964, **20**, 513–515.

BRANDT, L. W. Rejection of psychotherapy: The discovery of unexpected numbers of pseudo-rejectors. *Archives of General Psychiatry,* 1964, **10**, 310–312.

BRILL, N. Q., & STORROW, H. A. Social class and psychiatric treatment. *Archives of General Psychiatry,* 1960, **3**, 340–344.

BUDNER, S. S., ESCOVER, H., & MALITZ, S. The relationship of social personality, and psychiatric factors to choice of psychiatric therapy. *Comparative Psychiatry,* 1964, **5**, 327–333.

CHESS, S., CLARK, K. B., & THOMAS, A. Importance of cultural patterns in psychotherapy. *Psychiatric Quarterly,* 1953, **27**, 102–114.

CHESSICK, R. D. *Why psychotherapists fail.* New York: Science House, 1971.

COHEN, J. Social work and the culture of poverty. In F. Riessman, J. Cohen, & A. Pearl (Eds.), *Mental health of the poor.* New York: Free Press, 1964.

COLES, N. J., BRANCH, C. H. H., & ALLISON, R. B. Some relationships between social class and the practice of dynamic psychotherapy. *American Journal of Psychiatry,* 1962, **118**, 1004–1012.

DENGROVE, E., & KUTASH, S. B. Why patients discontinue treatment in a mental hygiene clinic. *American Journal of Psychotherapy,* 1950, **4**, 457–472.

FIERMAN, L. B. Myths in the practice of psychotherapy. *Archives of General Psychiatry,* 1965, **12**, 408–414.

FISHMAN, J. R. Poverty, race, and violence. *American Journal of Psychotherapy,* 1969, **23**, 599–607.

FISHMAN, J. R., & MCCORMACK, J. Mental health without walls: Programs for the ghettos. In J. H. Masserman (Ed.), *Current psychiatric therapies,* Vol. 9. New York: Grune & Stratton, 1969.

FRANK, J. D. *Persuasion and healing: A comparative study of psychotherapy.* New York: Schocken Books, 1961.

FRANK, J. D., GLIEDMAN, L. H., IMBER, S. D., NASH, E. H., & STONE, A. R. Why patients leave psychotherapy. *Archives of Neurology and Psychiatry,* 1957, **17**, 283–299.

GANS, H. J. *The urban villagers.* New York: Free Press, 1962.

GREEN, A. W. Social values and psychotherapy. *Journal of Personality,* 1946, **14**, 199–228.

GREY, A. L. Social class and the psychiatric patient: A study in composite character. In A. L. Grey, (Ed.), *Class and personality in society.* New York: Atherton, 1969.

GRIER, W. H., & COBBS, P. M. *Black rage.* New York: Bantam, 1968.

GUERNEY, B. *Psychotherapeutic agents: New roles for nonprofessionals, parents, and teachers.* New York: Holt, Rinehart & Winston, 1969.

GUNDLACH, R. H., & GELLER, M. The problem of early termination: Is it really the terminee? *Journal of Consulting Psychology,* 1958, **22**, 410.

HAGGSTROM, W. C. The power of the poor. In F. Riessman, J. Cohen, & A. Pearl (Eds.), *Mental health of the poor.* New York: Free Press, 1964.

HEINE, R. W., & TROSMAN, H. Initial expectations of the doctor-patient interaction as a factor in continuance in psychotherapy. *Psychiatry,* 1960, **23**, 275–278.

HOLLINGSHEAD, A. B. *Two-factor index of social position.* New Haven: Author, 1957.

HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social-class and mental illness.* New York: Wiley, 1958.

IMBER, S. D., NASH, E. H., & STONE, A. R. Social class and duration of psychotherapy. *Journal of Clinical Psychology,* 1955, **11**, 281–294.

IMBER, S. D., FRANK, J. D., GLIEDMAN, L. H., NASH, E. H., & STONE, A. R. Suggestibility, social class, and the acceptance of psychotherapy. *Journal of Clinical Psychology,* 1956, **12**, 341–344.

KADUSHIN, C. *Why people go to psychiatrists.* New York: Atherton, 1969.

KATZ, M. M., LORR, M., & RUBENSTEIN, E. A. Remainer patient attributes and their relation to subsequent improvement in psychotherapy. *Journal of Consulting Psychology,* 1958, **22**, 411–413.

KOEGLER, R. R., & BRILL, N. Q. *Treatment of psychiatric outpatients.* New York: Appleton-Century-Crofts, 1967.

LERNER, B. *Therapy in the ghetto.* Baltimore: John Hopkins University Press, 1972.

LORR, M., KATZ, M. M., & RUBENSTEIN, E. A. The prediction of length of stay in psychotherapy. *Journal of Consulting Psychology,* 1958, **22**, 320–327.

MCMAHON, J. T. The working class psychiatric patient: A clinical view. In F. Riessman, J. Cohen, & A. Pearl (Eds.), *Mental health of the poor.* New York: Free Press, 1964.

MELTZOFF, J., & KORNREICH, M. *Research in psychotherapy.* New York: Atherton, 1970.

MILLER, S. M. The American lower-classes: A typological approach. In F. Riessman, J. Cohen, & A. Pearl (Eds.), *Mental health of the poor.* New York: Free Press, 1964.

MILLER, S. M., & MISHLER, E. G. Social class, mental illness, and American psychiatry: An expository review. *Millbank Memorial Fund Quarterly,* 1959, **37**, 174–199.

MILLER, S. M., & RIESSMAN, F. The working-class subculture: A new view. In A. L. Grey (Ed.), *Class and personality in society.* New York: Atherton, 1969.

MINUCHIN, S., MONTALVO, B., GUERNEY, B. G., ROSMAN, B. L., & SCHUMER, F. *Families of the slums: An exploration of their structure and treatment.* New York: Basic Books, 1967.

MYERS, J. K., & AULD, F. Some variables related to outcome of psychotherapy. *Journal of Clinical Psychology,* 1955, **11**, 51–54.

MYERS, J. K., & BEAN, L. L. *A decade later: A follow-up of social class and mental illness.* New York: Wiley, 1968.

MYERS, J. K., & ROBERTS, B. H. *Family and class dynamics in mental illness.* New York: Wiley, 1959.

MYERS, J. K., & SCHAFFER, L. Social stratification and psychiatric practice: A study of an outpatient clinic. In S. K. Weinberg (Ed.), *The sociology of mental disorders: Analysis and readings in psychiatric sociology.* Chicago: Aldine, 1967.

NASH, E. H., HOEHN-SARIC, R., BATTLE, C. C., STONE, A. R., IMBER, S. D., & FRANK, J. D. Systematic preparation of patients for short-term psychotherapy II: Relation to characteristics of patient, therapist and the psychotherapeutic process. *Journal of Nervous and Mental Disease,* 1965, **140**, 374–383.

RAINWATER, L. Neutralizing the disinherited: Some psychological aspects of understanding the poor. In V. L. Allen, *Psychological factors in poverty.* Chicago: Markham, 1970.

RASKIN, A. Factors therapists associate with motivation to enter psychotherapy. *Journal of Clinical Psychology,* 1961, **17**, 62–65.

Redlich, F. C. Social aspects of psychotherapy. *American Journal of Psychiatry,* 1958, **114**, 800–805.

Redlich, F. C., Hollingshead, A. B., Roberts, B. H., Robinson, H. A., Friedman, L. Z., & Myers, J. K. Social structure and psychiatric disorders. *American Journal of Psychiatry,* 1953, **109**, 729–734.

Riess, B. F., & Brandt, L. W. What happens to applicants for psychotherapy? *Community Mental Health Journal,* 1965, **1**, 175–180.

Riessman, F., Cohen, J., & Pearl, A. *Mental health of the poor.* New York: Free Press, 1964.

Robinson, H. A., Redlich, F. C., & Myers, J. K. Social structure and psychiatric treatment. *American Journal of Orthopsychiatry,* 1954, **24**, 307–316.

Rosenthal, D., & Frank, J. D. The fate of psychiatric clinic outpatients assigned to psychotherapy. *Journal of Nervous and Mental Disorders,* 1958, **127**, 330–343.

Sanua, V. D. Sociocultural aspects of psychotherapy and treatment: A review of the literature. In L. E. Abt & B. F. Riess (Eds.), *Progress in clinical psychology.* New York: Grune & Stratton, 1966.

Sarbin, T. R. The culture of poverty, social identity, and cognitive outcomes. In V. L. Allen, *Psychological factors in poverty.* Chicago: Markham, 1970.

Savitz, H. A. The cultural background of the patient as part of the physician's armamentarium. *Journal of Abnormal and Social Psychology,* 1952, **47**, 245–254.

Schaffer, L., & Myers, J. K. Psychotherapy, and social stratification. *Psychiatry,* 1954, **17**, 83–93.

Schneiderman, L. Social class, diagnosis, and treatment. *American Journal of Orthopsychiatry,* 1965, **35**, 99–105.

Shader, R. I. The walk-in service: An experience in community care. In T. Rothman (Ed.), *Changing patterns in psychiatric care.* New York: Crown, 1970.

Silberman, C. E. *Crisis in black and white.* New York: Vintage, 1964.

Small, L. *The briefer psychotherapies.* New York: Brunner & Mazel, 1971.

Sobey, F. *The non-professional revolution in mental health.* New York: Columbia University Press, 1970.

Srole, L., Langer, T. S., Michael, S. T., Opler, M. K., & Rennie, T. A. C. *Mental health in the metropolis: The midtown Manhattan study.* New York: McGraw-Hill, 1962.

Strupp, H. H., & Bergin, A. E. Some empirical and conceptual bases for coordinated research in psychotherapy. *International Journal of Psychiatry,* 1969, **7**, 17–90.

Sullivan, P. L., Miller, C., & Smelser, W. Factors in length of stay and progress in psychotherapy. *Journal of Consulting Psychology,* 1958, **22**, 1–9.

Thomas, A., & Sillen, S. *Racism and psychiatry.* New York: Brunner & Mazel, 1972.

Truax, C. B., & Carkhuff, R. R. *Toward effective counseling and psychotherapy.* Chicago: Aldine, 1967.

Winder, A. E., & Hersho, M. The effect of social class in the length and type of psychotherapy in a Veterans' Administration Mental Hygiene Clinic. *Journal of Clinical Psychology,* 1955, **11**, 77–79.

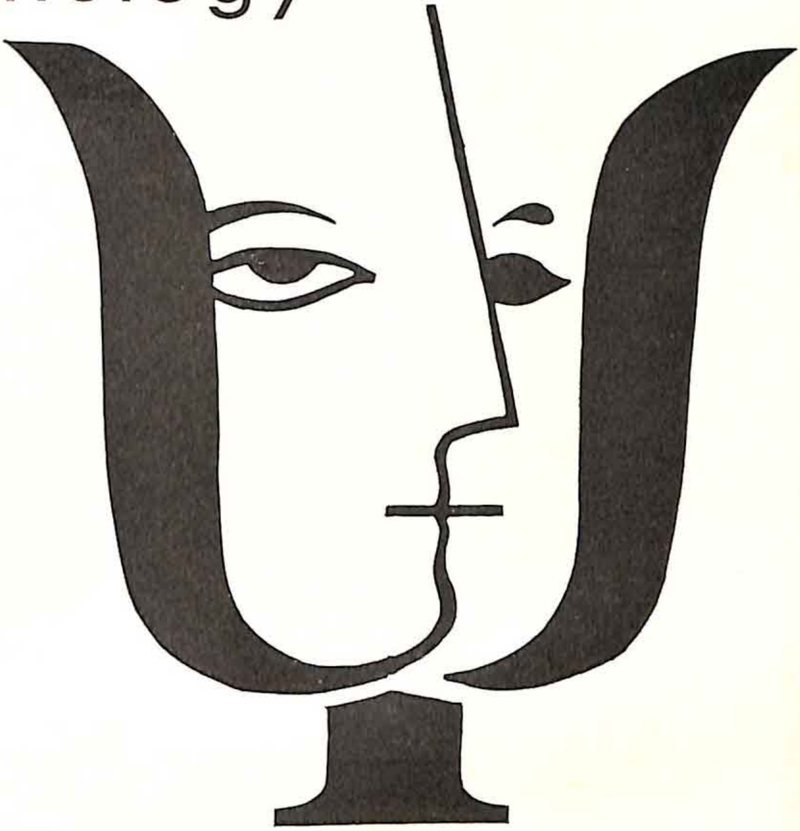
Yamamoto, J., & Goin, M. K. Social class factors relevant for psychiatric treatment. *Journal of Nervous and Mental Disease,* 1965, **142**, 332–339.

Yamamoto, J., James, Q. C., Bloombaum, M., & Hatter, J. Racial factors in patient selection. *American Journal of Psychiatry,* 1967, **124**, 630–636.

Zax, M., & Cowen, E. L. *Abnormal psychology: Changing conceptions.* New York: Holt, Rinehart & Winston, 1972.

(Received April 3, 1972)

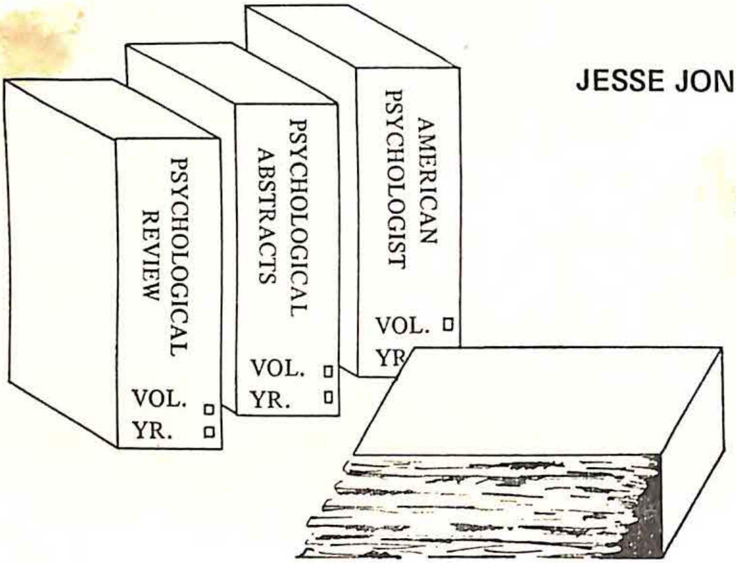
professional
psychology



Professional Psychology is devoted to the application of principles and skills in a wide variety of settings: clinics, hospitals, schools, industry, government, and independent practice. The journal contains original articles on both theory and techniques which will enhance knowledge and improve the effectiveness of psychologists. Quarterly: one volume per year. The 1973 volume is Vol. 4. Subscriptions: $12.00 (foreign $13.00).

American Psychological Association
1200 Seventeenth Street, N.W.
Washington, D.C. 20036

• Payment must accompany orders amounting to $15.00 or less •



JESSE JONES VOLUME FILES

Send orders to:
JESSE JONES CORP.
P.O. Box 5120
Philadelphia, Pa. 19141

Only $4.25 ea.
3 for $12.00
6 for $22.00 P.P.D.
USA Orders Only

Now, the journals of the American Psychological Association can become an attractive permanent part of your professional library. These famous Jesse Jones volume files, especially designed to keep your copies orderly, readily accessible for future references — guard against soiling, tearing, wear or misplacement of copies.

This durable file will support 150 lbs. Looks and feels like leather and is washable. The 23-carat gold lettering makes it a fit companion for the most costly binding.

Vol. 79, No. 5 May 1973

Psychological Bulletin

Copyright © 1973 by the American Psychological Association, Inc.

BEHAVIORAL DIFFERENCES BETWEEN ACUTE AND CHRONIC SCHIZOPHRENICS:

COURSE OF PSYCHOSIS, EFFECTS OF INSTITUTIONALIZATION, OR SAMPLING BIASES?[1]

MILTON E. STRAUSS[2]

Wayne State University

Scores of studies have been reported which demonstrate behavioral differences between acute and chronic schizophrenics. These data have most often been interpreted as reflecting changes in psychological functioning due to (*a*) processes intrinsic to the schizophrenic disorder and/or (*b*) the effects of institutionalization. Examination of the construct of chronicity and of the design of these studies suggests (*i*) the existence of perceptual and cognitive changes during the course of schizophrenia has not been demonstrated; (*ii*) the observed acute-chronic differences may primarily reflect sampling biases.

Behavioral variability within the broad psychiatric category of schizophrenia is well documented (Shakow, 1962, 1963; Silverman, 1964a). The subclassification of schizophrenic subjects in terms of three individual differences constructs often markedly reduces this variability, and significant differences among these schizophrenia subtypes are frequently observed (Broen, 1968; Silverman, 1964a). These dimensions are (*a*) adequacy of premorbid adjustment, (*b*) clinical subtype, specifically paranoid or not, and (*c*) chronicity. These dimensions have been used to classify subjects in scores of studies, sometimes in combination but most often singly (cf. Schooler & Feldman, 1967). These studies have demonstrated the empirical and conceptual value of the premorbid adjustment (or process-reactive) and paranoid symptomatology constructs for the development of an understanding of the structure of schizophrenia (Broen, 1968; Higgins, 1969; Garmezy, 1971; Shakow, 1962, 1963).

The chronicity construct, however, has less heuristic value. Its meaning is ambiguous and behavioral differences between schizophrenic groups classified in terms of chronicity are often inconsistent. Unusual word association responses, for example, have been reported as characteristic of chronic schizophrenics (Higgins, Mednick, & Philip, 1965), poor premorbid subjects (Dokecki, Polidoro, & Cromwell, 1965), more the latter than the former (Dokecki, Cromwell, & Polidoro, 1968), primarily the former (Higgins, 1968), and neither in any clear-cut way (Ries & Johnson, 1967). Other correlates of chronicity have been reported more consistently. In general, perceptual and cognitive measures suggest that chronic subjects use fewer cues and focus attention more narrowly than do acutes (cf. Broen, 1968). This finding has typically been interpreted as indicating that a progressive change occurs in the psychological functioning of schizophrenics. Two major hypotheses have been advanced to explain this apparent change.

Diary No 705(5)
4/1/74

---

[1] Preparation of this paper was supported in part by Grant No. MH21822-01 from the National Institute of Mental Health. Gratitude is expressed to Rue L. Cromwell for his extensive and constructive comments on earlier drafts of this paper and to the anonymous editorial consultant for his most helpful stylistic revisions and suggestions.

[2] Reprint requests should be sent to Milton E. Strauss, Department of Psychology, Wayne State University, Detroit, Michigan 48202.

The first proposes that acute schizophrenics, as they become chronic, progressively focus attention more narrowly in order to avoid the aversive psychological disorganization produced by normal ranges of cue utilization (e.g., Broen & Storms' [1967; Broen, 1968] response interference theory; Johannsen, Friedman, & Liccione, 1964; Mednick, 1967). The second hypothesis is that this narrowing of attention is a reaction to the impoverished, redundant environment of the hospital to which the chronics are exposed for long periods of time. This change then is not intrinsic to the course of the psychosis (Pishkin, Smith, & Liebowitz, 1962; Silverman, Berg, & Kantor, 1965).

Data consistent with both the course-of-psychosis and effects-of-institutionalization hypotheses have been reported (Johannsen & O'Connell, 1965; Silverman et al., 1965) but unequivocal conclusions are not possible. The only generalization which may be made from this rather large literature (cf. Schooler & Feldman, 1967) is that acute and chronic schizophrenics tend to differ on a number of behavioral measures. The reasons for this rather restricted conclusion are dealt with in the following sections.

RESEARCH DESIGN IN CHRONICITY STUDIES

Silverman (1964a, 1967), for one, has suggested that the variable of chronicity is crucial to the resolution of apparently inconsistent empirical findings and to the development of appropriate theories of perceptual, cognitive, and autonomic functioning in schizophrenia. The studies on which his analysis was based, and those subsequent to his reports (cf. Broen, 1968; Neale & Cromwell, 1970) have used cross-sectional designs. Common to these studies are (*a*) the sampling of an inpatient population at a single point in time; (*b*) the exclusion of subjects likely to introduce extraneous variance or to increase variability;[3] (*c*) the use of a time-bound metric, for example, length of illness, to classify subjects as acute or chronic. The two groups differ substantially in chronicity: The shortest time ill chronic subjects exceed the longest time ill acutes by one or more years. (*d*) The variables of premorbid adjustment and/or paranoid symptomatology are ignored in selecting subjects. While especially true of studies prior to Silverman's (1964a) review, most later studies sample only one of these two variables (e.g., Yates & Korboot, 1970). (*e*) Group differences on the dependent variable(s) are attributed to the selection principle explicitly employed in the composition of the groups, that is, chronicity.

A number of aspects of this paradigm limit the conclusions which may be drawn from the empirical findings. As with any static-group comparison (Campbell & Stanley, 1963), the attribution of the source of the behavioral difference is questionable. The groups may also differ on other variables whose relationships to the dependent behavioral variable are to varying degrees not known (McGhie, 1970). Further, as will be demonstrated below, acute and chronic groups do differ on variables of known importance: life history and symptom characteristics. Studies could be conducted to detect potentially confounding relationships and their possible effects could be "controlled" by means of subject selection or statistical techniques. This would be of limited utility, however, for the fundamental source of the uncertainty of the meaning of behavioral differences between acute and chronic subjects is the use of static-group designs to examine developmental hypotheses. The inappropriateness of these designs may be best appreciated by the consideration of typical definitions of chronicity.

CONSTRUCT OF CHRONICITY

Within the time-linked conception employed by most psychological investigators, chronicity refers to either length of time elapsed since the onset of schizophrenia, that is, length of illness, or to the total length of hospitalization since onset (Cromwell, 1972).[4] In both cases, onset is defined as the time of first ad mission to a hospital, an empirically useful but far from ideal criterion (Neale & Cromwell, 1970).

---

[3] These typically include organic brain disease, alcoholism, recent electroconvulsive therapy and uncooperativeness. The biases introduced by these exclusions, especially the last, are not unique to this area (cf. Chapman, 1963; Klein & Spohn, 1964).

[4] It is unclear in the literature whether length o hospitalization refers to only continuous hospitaliza tions, total hospitalizations, or both.

An acute, or more accurately, early-term schizophrenic then is (*a*) a recent first admission, (*b*) a first admission with only a brief hospitalization at the time of the study, or (*c*) a multiple-admissions patient with a small total length of hospitalization. Chronic or long-term subjects are those who (*a*) were first hospitalized a substantial time prior to the study and who may or may not have substantial length of hospitalization or (*b*) have had long length of hospitalization. Although investigators have used the several meanings and measures of chronicity interchangeably, their empirical equivalence has not been demonstrated.

Before the widespread introduction of pharmacological and other treatment changes in the mid-1950s, length of illness and length of hospitalization were very likely substantially correlated. This is suggested by Brown's (1960) summary of epidemiological survey of patients admitted between 1900 and 1951. For first admission cohorts, there was a negative linear relationship between probability of discharge and length of continuous hospitalization during the first two years after admission. In the third year of hospitalization, discharge rates plummeted markedly and became stabilized at a low level after three years length of hospitalization. During this time, most hospitalized schizophrenics were first admissions and computing chronicity in terms of months since first admission or months of hospitalization since first admission would lead to almost identical numerical values. However, the treatment careers of schizophrenics have changed markedly since the time period considered in Brown's study (Smith, Bower, & Wignall, 1965; Yolles & Kramer, 1969). Intermittent patienthood (Friedman, von Mering, & Hinko, 1966), the more typical life course of schizophrenics currently, has almost certainly attenuated the correlation between length of illness and length of hospitalization. This may account for the differences in findings between procedurally similar studies which define chronicity differently (e.g., Kopfstein & Neale, 1971; Strauss & Parwatikar[5]). Further, when studies which define chronicity differently show convergences in findings, their interpretation still remains unclear. Long-term hospitalized subjects and long-term ill but only intermittently hospitalized subjects differ in significant ways. The latter become rehospitalized as a function of psychosocial characteristics and their failure to continue taking ataractic drugs (Crumpton, 1968; Paul, 1969; Rosen, Englehardt, Freedman, & Margolis, 1968). The continuously hospitalized subjects, among other things, have not responded to phenothiazines more than minimally.

Most studies treat chronicity as a typology, although it is measured dimensionally. Hence it becomes important to select an appropriate cutting point for discriminating between early-term and long-term subjects. On the basis of Brown's (1960) work, it has been recommended that two years length of hospitalization be used as the maximum for inclusion in the early-term group, and at least three years length of hospitalization be used as the minimum for the long-term group. Subjects in the more ambiguous two to three year length of hospitalization range are excluded (Neale & Cromwell, 1970; Neale, Held, & Cromwell, 1969). The current validity of Brown's (1960) conclusions may be questioned. Although the data are not as extensive for the post-1950s, it appears that six months or one year of continuous hospitalization are currently more accurate inflection points in probability of discharge curves (Gorwitz, Bahn, Klee, & Solomon, 1966; Yolles & Kramer, 1969). In the absence of appropriate epidemiological data for current first admission and resident populations, it is problematic to determine a suitable length of hospitalization criterion for the identification of chronicity.

Silverman's (1964a) alternative criteria (early term < 3 years length of hospitalization, long term > 6 years length of hospitalization) were based on his analysis of studies which suggested that these criteria identified patient subgroups which were quite different on perceptual and cognitive measures. These groups also differed in other important ways

---

[5] M. E. Strauss & S. Parwatikar. Size estimation in chronic schizophrenia. Paper presented at the meeting of the Eastern Psychological Association, New York, April, 1971.

(e.g., phenothiazine treatment status, premormid adjustment, and paranoid status) and the theoretical propositions associated with these chronicity criteria have subsequently had a mixed validity history (Kopfstein & Neale, 1971; Neale & Cromwell, 1970; Neale, Davis, & Cromwell, 1971). The meaningfulness of these criteria are thus also uncertain.

The use of cutting points to partition a continuous variable into discrete categories affects the likelihood of detecting relationships with other variables and the nature of the effects detected. If length of illness or length of hospitalization measures are arbitrarily dichotomized into early-term and long-term categories, the statistical power of studies is reduced; but if subjects of intermediate chronicity (e.g., $3 <$ length of hospitalization < 6 years) are discarded, the likelihood of finding effects may be increased. In either case the use of dichotomous chronicity measures precludes the detection of linear or curvilinear relationships with dependent variables. Such relationships are easily detectable when appropriate correlational techniques are applied to a continuous measure of chronicity like length of illness or length of hospitalization.

Length of hospitalization appears to be the preferred measure of chronicity in the recent literature. Like length of illness, length of hospitalization is an easily collected datum; it is obtained from hospital records. However, for subjects with multiple admissions at several hospitals, length of illness is more easily and reliably ascertained as only the date of first admission need be determined. The preference for length of hospitalization would appear to have three sources (*a*) the historical role of length of hospitalization and discharge criteria as measures of prognosis and outcome in schizophrenia (Brown, 1960; Yolles & Kramer, 1969); (*b*) a belief that institutionalizational itself produces deterioration in mental patients (cf. Williams[6]); and (*c*) a provocative report by Silverman et al. (1965).

In this oft-cited study it was found that early-term and long-term institutionalized felons differed on the Titchener Circles illusion and on the Pettigrew Category Width tasks in the same way that early-term and long-term schizophrenics were expected to differ. This appeared to demonstrate effects of institutionalization independently of schizophrenic psychopathology. However, in the study of schizophrenics to which the prisoners' data were compared (Silverman, 1964b) significant early-term–long-term differences were found only for paranoids, were post hoc, and were clear only on the task for which predicted paranoid–nonparanoid differences were not found. At best, these unreplicated findings are applicable to only some schizophrenics—the paranoids. Further, in a study using well-established reaction time and conceptual measures Best[7] found early-term–long-term differences for schizophrenics but not for matched prisoner controls.

Johannsen and O'Connell (1965) directly questioned the importance of the institutionalization variable in a reanalysis of an earlier study (Johannsen et al., 1964). In the latter study, in which chronicity was treated as the continuous variable which it is, a number of visual perception measures were found to be related to length of illness. With the exception of the length of illness < 6 months group, all subjects had spent equivalent proportions of time since first admission in hospitals. If institutionalization itself were an important determinant of schizophrenics' behavior, they reasoned, high proportion of time and low proportion of time hospitalized subjects should differ on perceptual tasks.[8] Only one significant difference was found in contrast with a number of differences associated with length of illness.

Johannsen's work has been cited as a basis for selecting early-term and long-term subjects using length of illness rather than length of hospitalization (Neale & Cromwell, 1970).

---

[6] P. H. Williams. *The prevention of disability in mental disorders*. Washington, D. C.: Department of Health, Education and Welfare, 1962.

[7] R. H. Best. Cognitive impairment in schizophrenics and prison inmates as a function of prolonged exposure to impoverished environmental conditions. (Doctoral dissertation, Southern Illinois University) Ann Arbor, Mich.: University Microfilms 1968, No. 69-1739.

[8] This measure is similar to the proportion of lifetime hospitalized index used in DeWolfe's (1971) research program. This measure is highly correlated ($r = .99$) with length of hospitalization (DeWolfe personal communication, March 30, 1972).

Given the vagaries of discharge and readmission policies (Cumming, 1963; Mosher & Feinsilver[9]), the use of length of illness would appear to have much to recommend itself. Of course, if only two groups of subjects are selected, the problem of appropriate cutting points still remains. However, further examination of the Johannsen study suggests that the length of illness findings may reflect the operation of a subject sampling variable known to be related to behavioral variation in schizophrenia-paranoid versus nonparanoid symptomatology. A secondary analysis by this writer revealed a significant linear decrease in the proportion of paranoid subjects as length of illness increased (linear trend $X^2 = 6.5171$, $df = 1$, $p < .02$; residual X^2, ns). A similar trend was found in a secondary analysis of Sanes and Zigler's (1971) recent study of the relationship between premorbid adjustment and paranoid symptomatology. When this writer divided their small sample ($N = 37$) into first-admission and multiply-admitted groups, there was a significant trend ($X^2 = 4.5596$, $df = 1$, $p < .05$) for the latter to be less often paranoid.

An hypothesis other than sampling bias can also account for these findings. Perhaps older subjects are less often paranoid than younger subjects. Also, paranoid symptoms may have disappeared over the course of the psychosis and the repeated hospitalization experienced by Johannsen's subjects (Morgan, Porzio, & Hedlund, 1968). Similarly, Sanes and Zigler's multiply-admitted subjects may have been paranoid at first admission.[10] The crucial issue is: Do the paranoid patients or the paranoid symptoms disappear from chronic or long-term populations? This is a question to which we shall return.

Suggestions of a sampling bias hypothesis (or sample-composition-changes hypothesis) to account for early-term–long-term differences may be found in the literature. Payne (1966), for example, observed that early-term subjects were either overinclusive or retarded while long-term subjects were primarily retarded, and speculated that overinclusion might be a good prognostic sign. Hence early-term–long-term differences on measures of overinclusion could reflect the loss of overinclusive subjects from long-term groups rather than any cognitive change (cf. also Draguns, 1963). A small sample follow-up study of mental patients, all of whom were not schizophrenic, provided some tentative support for this hypothesis (Payne, 1968). The changing-sample-composition hypothesis is thus not a new one. It is, however, an unexplored one and has substantial significance for the interpretation of cross-sectional comparison tests of the developmental course-of-psychosis and effects-of-institutionalization hypotheses.

SOURCES OF SAMPLING BIAS IN STUDIES COMPARING EARLY-TERM WITH LONG-TERM SCHIZOPHRENICS

Two sources of sampling bias in cross-sectional studies of early-term and long-term schizophrenics (premorbid adjustment and paranoid symptomatology) are parameters of substantial import in the description of variability in schizophrenics' behavior and are dealt with first. Subsequently, this section deals with other sources of sampling bias in these designs.

Premorbid Adjustment

Studies of early-term and long-term schizophrenics must sample institutionalized populations if chronicity is assessed using length of hospitalization. Although the use of length of illness measures does not require the sampling of only inpatient populations, studies have been restricted to them. It is well documented that premorbid adjustment predicts behavioral differences among early-term schizophrenics (Higgins, 1969). In general, poor premorbid subjects manifest at first admission greater degrees of psychological deficit than do good premorbid subjects (Higgins, 1969; Neale and Cromwell, 1970). Good premorbid subjects have substantially higher probability of discharge, significantly shorter durations of hospitalization at admission or readmission, and are less likely to be readmitted to inpatient facilities (Farina & Webb, 1956; Rosen, Klein, & Gittleman-Klein, 1971; Turner & Zabo, 1968). Thus good premorbids are underrepresented in long-term samples. Parsimony would suggest that early-term–long-term differences be interpreted in terms of differences in premorbid adjustment, rather than in terms of differences or changes associated with chronicity.

---

[9] L. Mosher & D. Feinsilver. *Special Report: Schizophrenia*. (NIMH-HSM Tech. Rep. No. 72-9007) Washington, D. C.: United States Government Printing Office, 1971.

[10] Paranoid symptoms and premorbid adjustment vere unrelated in this study. Hence this trend can-ot be accounted for by a differential incidence of ood and poor premorbid subjects in the two num-ers of admission groups. Premorbid adjustment data ere not provided by Johannsenn et al. (1964).

Investigators aware of the premorbid adjustment bias in randomly selected long-term subjects have suggested a matching strategy (Neale & Cromwell, 1970). Selecting subjects in terms of premorbid adjustment as well as chronicity criteria would assure the numerical equivalence of the proportion of good premorbids in early-term and long-term groups. This procedure involves the tacit assumption that the less frequently found long-term goods are a representative sample of good premorbid subjects. The good premorbid long-term schizophrenic is deviant with respect to outcome, however. Is it not likely that he was also deviant behaviorally at admission or during the early-term period? The answer to this question is not known; behavioral studies of early-term good premorbids who experience different long-term outcomes could not be found.

Paranoid Symptomatology

It has been pointed out that paranoid subjects are less often found in chronic groups. Among early-term subjects, paranoids tend to manifest less psychological deficit than nonparanoid schizophrenics (Broen, 1968). Whether the sampling bias position applied to the premorbid adjustment variable is also applicable here depends on the source of the underrepresentation of paranoids in long-term groups. As was asked earlier, do paranoid patients or paranoid symptoms disappear?

The answer to this question is uncertain. Clinical lore supports either view (Bleuler, 1950; Neale and Cromwell, 1970; Robins and Guze, 1970). It is clear that among readmitted patients fewer manifest paranoid symptoms on second admission (e.g., Morgan et al., 1968), but these studies do not indicate what proportion of paranoid and nonparanoid first admissions are rehospitalized.

There are suggestive but nonconclusive findings supporting the hypothesis that paranoid patients disappear from hospitalized samples. First, Sommer and Witney (1961), discussing the "chain of chronicity," noted that twice as many nonparanoid as paranoid patients were transferred from short-term intensive treatment to chronic care wards. Paranoids were more likely to be directly discharged. The second source of information was a follow-up of Stenn's [11] Veterans Administration subjects. Three years after the initial study of these early-term (length of hospitalization < 3 years) subjects, we examined their rates of discharge and readmission. Of the 29 paranoid subjects whose records were located, 45% were not in hospitals at any point during the 3-year follow-up, but 73% of the nonparanoids had been either readmitted or not discharged during the same span of time ($z = 2.0737$, $p < .05$). This difference in the incidence of paranoids in what are now long-term subjects supports the loss-of-patients hypothesis. This conclusion must be viewed as tentative as the numbers are small and it is possible, although unlikely, that the paranoids were rehospitalized in other than Veterans Administration settings. It should also be noted that the loss of paranoid symptoms hypothesis may also be valid; the initially paranoid subjects who were rehospitalized may have been nonparanoid at readmission. Unfortunately, readmission diagnoses were not available.

Intermittent Patienthood

Baldly stated, samples of hospitalized long-term subjects are extremely unrepresentative of the long-term schizophrenic population whether defined by length of illness or length of hospitalization criteria. Drug treatment has markedly facilitated the development of treatment programs for schizophrenic patients outside of hospitals (Pasamanick, Scarpitti, &

---

[11] P. G. Stenn. The relation between eye movement frequency and size estimation in schizophrenics and nonpsychiatric patients. Paper presented at the meeting of the Eastern Psychological Association Philadelphia, 1969. Appreciation is extended to P. G Stenn for making the original data available and D. Bidus, L. Brown, and J. Basham of the St. Loui Missouri, Cincinnati, Ohio, and Lexington, Kentuck Veterans Administration Hospitals for facilitating th patient follow-up.

Dinitz, 1967). A major determinant of whether a patient will become rehospitalized, and hence be a potential subject, is whether prescribed ataractic drugs continue to be taken (Crumpton, 1968). Rehospitalized subjects are likely to be those who have voluntarily discontinued drug regimens and it appears that drug discontinuance results in rehospitalization for essentially only poor premorbid subjects. The use of a placebo instead of an active drug does not affect the hospitalization rates of "nonhospitalization prone" (i.e., good premorbid) subjects (Freedman, Rosen, Englehardt, & Margolis, 1967; Rosen et al., 1968).

Other factors contribute to nonrepresentativeness of hospitalized long-term schizophrenics and the uninterpretability of early-term–long-term differences: Rehospitalized or long-term hospitalized subjects are not invariably symptomatic. Readmission may depend less on degree of psychopathology or on an increase in symptomatic behavior than on posthospital living arrangements and other social variables (Brown, Carstairs, & Topping, 1958; Cumming, 1963; Mendel & Rappaport, 1969). Whether cognitive deficits are manifested may depend less on symptomatic status than on educational attainment and motivation for discharge (Braginsky, Braginsky, & Ring, 1969; Schwartz, 1967). The rapidly developing and widely implemented nonhospital treatment programs virtually guarantee nonrepresentativeness in resident long-term schizophrenic samples.

Phenothiazine Drug Treatment

The widespread use of phenothiazine and related drugs is problematic for behavioral studies of schizophrenia generally (Chapman, 1963; Lang & Buss, 1965). Recent studies indicate substantial heterogeneity in first admission and readmitted schizophrenics' behavioral and psychophysiological responses to ataractic drugs (Goldberg, 1970; Goldstein, 1970). Both in and out of the hospital, it is he poor premorbids who benefit from drugs. hese findings suggest that in early-term–ng-term comparisons group differences will e a joint function of drug treatment and subct selection procedures.

Consider a hypothetical study comparing cognitive functioning in early-term and long-term drug-treated, hospitalized, poor premorbid subjects. Will differences be found? Probably not if long-term subjects are selected from recently admitted patient populations for both early-term and long-term groups would be responsive to phenothiazine treatment. If, however, long-term subjects were selected from populations characterized by long, uninterrupted hospitalizations, differences would probably be observed. One would be comparing a maximally drug responsive group, poor early terms, with the residual, nonresponsive members of a poor premorbid cohort, the long terms. The observed early-term–long-term differences in the latter case, however, may reflect differences between drug-responsive and nonresponsive poors which existed during the early-term period.

CONCLUSIONS

Despite the arbitrary cutting points used to differentiate early-term and long-term schizophrenics, cross-sectional studies have revealed behavioral differences between these groups. Comparisons of hospitalized early-term and long-term subjects, however, are comparisons of differently heterogeneous groups: paranoid and nonparanoid, good and poor premorbid, drug responsive and nonresponsive subjects are compared with primarily nonparanoid, primarily poor premorbid, drug responsive, and nonresponsive subjects. Early-term–long-term differences are also affected by the selective discharge, retention, and readmission of schizophrenics associated with social, psychological, and psychopharmacological variables. Thus the conclusion that these behavioral differences reflect changes in psychological functioning is not warranted by the data. They are equally consistent with a more parsimonious changing-sample-composition hypothesis which does not require the inference of *intra*individual change to account for *inter*-individual differences. The study of changes in psychological processes during the course of this psychosis requires the longitudinal investigation of first admission cohorts, not the cross-sectional comparison of inpatient samples.

REFERENCES

Bleuler, E. *Dementia praecox or the group of schizophrenias.* (Trans. by J. Zinkin) New York: International Universities Press, 1950.

Braginsky, E. M., Braginsky, D. D., & Ring, K. *Methods of madness: The mental hospital as a last resort.* New York: Holt, Rinehart & Winston, 1969.

Broen, W. E., Jr. *Schizophrenia: Research and theory.* New York: Academic Press, 1968.

Broen, W. E., Jr., & Storms, L. H. A theory of response interference in schizophrenia. In B. A. Maher (Ed.), *Progress in experimental personality research.* Vol. 4. New York: Academic Press, 1967.

Brown, G. W. Length of hospital stay and schizophrenia: A review of statistical studies. *Acta Psychiatrica et Neurologica Scandinavica,* 1960, **35**, 414–430.

Brown, G. W., Carstairs, G. M., & Topping, G. Post-hospital adjustment of chronic mental patients. *Lancet,* 1958, 2, 685–689.

Campbell, D. T., & Stanley, J. C. *Experimental and quasi-experimental designs for research.* Chicago: Rand-McNally, 1963.

Chapman, L. J. The problem of selecting drug-free schizophrenics for research. *Journal of Consulting Psychology,* 1963, **27**, 540–542.

Cromwell, R. L. Strategies for studying schizophrenic behavior. *Psychopharmacologia,* 1972, **24**, 121–146.

Crumpton, N. E. The role of drugs in maintaining patients in the community. In G. Daniel & H. Freeman (Eds.), *The treatment of mental disorders in the community.* Baltimore: Williams & Wilkins, 1968.

Cumming, J. The inadequacy syndrome. *Psychiatric Quarterly,* 1963, **27**, 723–733.

DeWolfe, A. S. Cognitive structure and pathology in associations of process and reactive schizophrenics. *Journal of Abnormal Psychology,* 1971, **78**, 148–153.

Dokecki, P. R., Polidoro, L. G., & Cromwell, R. L. Commonality and stability of word association responses in good and poor premorbid schizophrenics. *Journal of Abnormal Psychology,* 1965, **70**, 312–316.

Dokecki, P. R., Cromwell, R. L., & Polidoro, L. G. The premorbid adjustment and chronicity dimensions as they relate to commonality and stability of word association in schizophrenics. *Journal of Nervous and Mental Disease,* 1968, **146**, 310–311.

Draguns, J. G. Responses to cognitive and perceptual ambiguity in chronic and acute schizophrenics. *Journal of Abnormal and Social Psychology,* 1963, **66**, 24–30.

Farina, A., & Webb, W. W. Premorbid adjustment and subsequent discharge. *Journal of Nervous and Mental Disease,* 1956, **124**, 612–613.

Freedman, N., Rosen, B., Engelhardt, D. M., & Margolis, R. Prediction of psychiatric hospitalization: 1. The measurement of hospitalization proneness. *Journal of Abnormal Psychology,* 1967, **72**, 468–477.

Friedman, I., von Mering, O., & Hinko, E. N. Intermittent patienthood: The hospital career of today's mental patient. *Archives of General Psychiatry,* 1966, **14**, 386–392.

Garmezy, H. Commentary. *Journal of Nervous and Mental Disease,* 1971, **153**, 317–322.

Goldberg, S. C. Drug effects and life history in schizophrenia. In M. Roff & D. F. Ricks (Eds.), *Life history research in psychopathology.* Minneapolis: University of Minnesota Press, 1970.

Goldstein, M. J. Premorbid adjustment, paranoid status, and patterns of response to phenothiazine in acute schizophrenia. *Schizophrenia Bulletin,* 1970, No. 3, 24–37.

Gorwitz, K., Bahn, A. K., Klee, G., & Solomon, M. Release and return rates for patients in state mental hospitals of Maryland. *Public Health Reports,* 1966, **81**, 1095–1108.

Higgins, J. Communality of word association responses in schizophrenia as a function of chronicity and adjustment: A response to Dokecki, Cromwell and Polidoro. *Journal of Nervous and Mental Disease,* 1968, **146**, 312–313.

Higgins, J. Process-reactive schizophrenia: Recent developments. *Journal of Nervous and Mental Disease,* 1969, **149**, 450–472.

Higgins, J., Mednick, S. A., & Philip, F. J. Associative disturbance as a function of chronicity in schizophrenia. *Journal of Abnormal Psychology,* 1965, **70**, 451–452.

Johannsen, W. J., Friedman, S. H., & Liccione, J. V. Visual perception as a function of chronicity in schizophrenia. *British Journal of Psychiatry,* 1964, **110**, 561–570.

Johannsen, W. J., & O'Connell, M. J. Institutionalization and perceptual decrement in chronic schizophrenia. *Perceptual and Motor Skills,* 1965, **21**, 244–246.

Klein, E., & Spohn, H. Further comments on characteristics of untestable chronic schizophrenics. *Journal of Abnormal and Social Psychology,* 1964, **68**, 355–358.

Kopfstein, J. H., & Neale, J. M. Size estimation in schizophrenic and non-schizophrenic subjects. *Journal of Consulting Psychology,* 1971, **36**, 430–435.

Lang, P. J., & Buss, A. H. Psychological deficit in schizophrenia: II. Interference and activation. *Journal of Abnormal Psychology,* 1965, **70**, 77–106.

McGhie, A. Attention and perception in schizophrenia. In B. A. Maher (Ed.), *Progress in experimental personality research.* Vol. 5. New York: Academic Press, 1970.

Mednick, S. A. The children of schizophrenics: Serious difficulties in current research methodologie which suggest the use of a "high-risk group method. In J. Romano (Ed.), *The origins schizophrenia.* New York: Excerpta Medica, 196

Mendel, W. M., & Rappaport, S. Determinants the decision for psychiatric hospitalization. *A chives of General Psychiatry,* 1969, **20**, 321–328.

MORGAN, D. W., PORZIO, R. J., & HEDLUND, J. L. Schizophrenic symptom change with rehospitalization. *Archives of General Psychiatry*, 1968, **19**, 227–231.

NEALE, J. M., & CROMWELL, R. L. Attention and schizophrenia. In B. A. Maher (Ed.), *Progress in experimental personality research*. Vol. 5. New York: Academic Press, 1970.

NEALE, J. M., HELD, J. R., & CROMWELL, R. L. Size estimation in schizophrenics: A review and reanalysis. *Psychological Bulletin*, 1969, **71**, 210–221.

NEALE, J. M., DAVIS, D., & CROMWELL, R. L. Size estimation in schizophrenia: Some additional controls. *Perceptual and Motor Skills*, 1971, **32**, 363–367.

PASAMANICK, B., SCARPITTI, F. R., & DINITZ, S. *Schizophrenics in the community*. New York: Appleton-Century-Crofts, 1967.

PAUL, G. L. Chronic mental patient: Current status—future directions. *Psychological Bulletin*, 1969, **71**, 81–94.

PAYNE, R. W. Long-term prognostic implications of overinclusive thinking in mental patients: A follow-up study using objective tests. *Proceedings of the IV World Congress of Psychiatry, Madrid, 1966*. New York: Experta Medica Foundation, 1968.

PAYNE, R. W. The measurement and significance of overinclusive thinking and retardation in schizophrenic patients. In P. H. Hoch & J. Zubin (Eds.), *Psychopathology of schizophrenia*. New York: Grune & Stratton, 1966.

PISHKIN, V., SMITH, T. E., & LIEBOWITZ, H. W. The influence of symbolic stimulus value on perceived size in chronic schizophrenia. *Journal of Consulting Psychology*, 1962, **26**, 323–330.

RIES, H. A., & JOHNSON, M. H. Commonality of word associations and good and premorbid schizophrenia. *Journal of Abnormal Psychology*, 1967, **72**, 487–488.

ROBINS, E., & GUZE, S. B. Establishment of diagnostic validity in psychiatric illness: Its application to schizophrenia. *American Journal of Psychiatry*, 1970, **126**, 983–987.

ROSEN, B., KLEIN, D. F., & GITTLEMAN-KLEIN, R. The prediction of rehospitalization: The relationship between age of first psychiatric treatment contact, marital status and premorbid asocial adjustment. *Journal of Nervous and Mental Disease*, 1971, **152**, 17–22.

ROSEN, B., ENGELHARDT, D. M., FREEDMAN, N., & MARGOLIS, R. The hospitalization proneness scale as a predictor of response to phenothiazine treatment: II. Prevention of psychiatric hospitalization. *Journal of Nervous and Mental Disease*, 1968, **146**, 476–480.

SANES, J., & ZIGLER, P. Premorbid social competence in schizophrenia. *Journal of Abnormal Psychology*, 1971, **78**, 140–144.

SCHOOLER, C., & FELDMAN, S. E. (Eds.) *Experimental studies of schizophrenia*. Goleta, Calif.: Psychonomic Press, 1967.

SCHWARTZ, S. Cognitive deficit among remitted schizophrenics: The role of a life-history variable. *Journal of Abnormal Psychology*, 1967, **72**, 54–58.

SHAKOW, D. Segmental set: A theory of the formal psychological deficit in schizophrenia. *Archives of General Psychiatry*, 1962, **6**, 1–17.

SHAKOW, D. Psychological deficit in schizophrenia. *Behavioral Science*, 1963, **8**, 275–305.

SILVERMAN, J. The problem of attention in research and theory in schizophrenia. *Psychological Review*, 1964, **71**, 352–379 (a).

SILVERMAN, J. Scanning control mechanism and "cognitive filtering" in paranoid and nonparanoid schizophrenia. *Journal of Consulting Psychology*, 1964, **28**, 385–393 (b).

SILVERMAN, J. Variations in cognitive control and psychophysiological defense in the schizophrenias. *Psychosomatic Medicine*, 1967, **29**, 225–245.

SILVERMAN, J., BERG, P. S. D., & KANTOR, R. Some perceptual correlates of institutionalization. *Journal of Nervous and Mental Disease*, 1965, **141**, 651–657.

SMITH, T. C., BOWER, W. H., & WIGNALL, C. M. Influence of policy and drugs on Colorado State Hospital population. *Archives of General Psychiatry*, 1965, **12**, 352–362.

SOMMER, R., & WITNEY, G. The chain of chronicity. *American Journal of Psychiatry*, 1961, **118**, 111–117.

TURNER, R. J., & ZABO, L. J. Social competence and schizophrenic outcome: An investigation and critique. *Journal of Health and Social Behavior*, 1968, **1**, 41–51.

YATES, A., & KORBOOT, P. Speed of perceptual functioning in chronic nonparanoid schizophrenics. *Journal of Abnormal Psychology*, 1970, **76**, 453–461.

YOLLES, S. F., & KRAMER, M. *Vital statistics*. In L. Bellak & L. Loeb (Eds.), *The schizophrenic syndrome*. New York: Grune & Stratton, 1969.

(Received June 5, 1972)

Psychological Bulletin
1973, Vol. 79, No. 5, 280–293

POWER RELATIONSHIPS AND GROUP DECISION MAKING IN ORGANIZATIONS

MICHAEL T. WOOD[1]

Ohio State University

Theory and research on the distribution of power in organizations are reviewed. Theories stressing power centralization are contrasted with the contemporary emphasis on the sharing of decision-making power. A situational-process approach to power relationships and group decision making integrates previous research and directs future study. This approach views power relationships as intervening processes in organizational development, with multiple determinants and outcomes. The group decision process is characterized as a multiphased process, in which participation, mutiple bases of power, and interaction dynamics affect power relationships. Linkages between decision processes and power, and between power and organizational consequences, are mediated by characteristics of the organizational environment, decision tasks, and individual motives.

"One overriding impression conveyed by surveying the literature of the 1960s, in contrast to the preceding two decades, is the redirection of interest in leadership processes such as power . . . relationships [Hollander & Julian, 1969, p. 388]." This paper contrasts theoretical emphases on power centralization and power sharing and presents an integrative model focusing on interactive organizational process-situation determinants of power relationships and organizational behavior. The model is presented in Figure 1.

The general approach it represents has two basic elements. First, organizational development is characterized as a process system (Katz & Kahn, 1966) with inputs generating outputs through the mediation of psychological processes and with these outputs feeding back into the system to affect further inputs. Second, the relationships between the three classes of parameters are mediated by individual and situational differences. The major parameters of the model to be examined in this paper are the input of the group decision-making process, intervening power relationships, outputs of decision characteristics and organizational effectiveness, and mediating variables associated with the organizational context, decision-making tasks, and individual differences in motivation. Previous models of personnel decisions (Dunnette, 1966), and leadership effectiveness (Campbell, Dunnette, Lawler, & Weick, 1970; Fiedler, 1967) offer precedents for a situational-process approach.

THEORETICAL BASES OF POWER SHARING AND GROUP DECISION MAKING

Early writings on the management of business and industrial organization emphasized the centralization of power in top management levels. Given the assumptions that the dominant worker motivation toward security could be fulfilled through economic reward, and that organizational efficiency was maximized through job simplification, centralization was in the mutual interest of the worker and the firm (Taylor, 1947). Functional viewpoints dictated specialists in planning and coordination (Gulick, 1937), and principles of management and control were oriented toward formal, hierarchical organization structures. Fayol (1949) argued for unity of command and scalar power chains; employees should receive orders from only one superior who controls homogeneous sets of activities, and excepting emergency situations, communications must travel up the hierarchy and be relayed through management down to the lower levels.

Bureaucratic theory (Weber, 1960) vested power in management positions structured by formal rules and distinct authority gradations. Group decision making was incompatible with

---

[1] Requests for reprints should be addressed to Michael T. Wood, Department of Psychology, Ohio State University, 404-C West 17th Avenue, Columbus, Ohio, 43210.

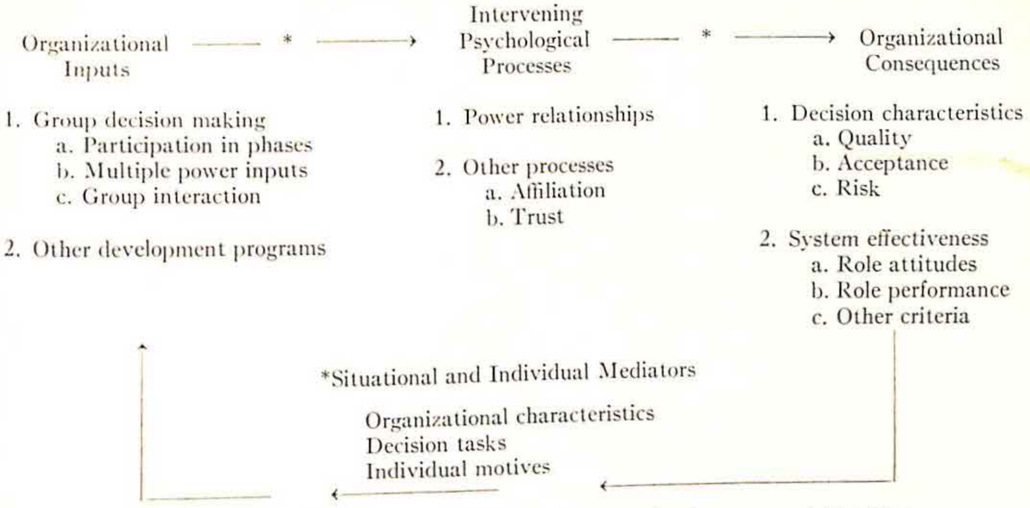


FIG. 1. A situational-process approach to organizational power relationships.

these structural positions. Planning boards and committees were implemented only in special circumstances since they violated the unity of command principle and brought people from different authority levels together to render decisions. These administrative theories of organization and their implied power relationships lacked empirical support. Today they are often treated as "straw men" (Bennis, 1970) or their assumptions are greatly modified (Gouldner, 1954; Merton, 1963; Udy, 1966).

Many organization theorists have abandoned these classical assumptions in favor of a power-sharing thesis, which rests on a multifaceted theoretical base incorporating complex motivational patterns and group processes. These theorists imply or directly suggest that group decision making can enhance organizational effectiveness through the fulfillment of social and ego motives, which they presume to be salient in the work setting.

Maslow's Need Hierarchy

Maslow's (1954) "need hierarchy" concept challenged the primary role of economic motivation. This position assumes that (*a*) man works in order to satisfy needs; (*b*) needs are hierarchically prepotent and categorized from lowest to highest level as physiological, safety, social, ego, and self-actualization; (*c*) satisfied needs do not motivate behavior; when a lower level need is satisfied, the immediately higher level of needs becomes the prime motivating force. Power centralization is incompatible with the need hierarchy hypothesis. "The philosophy of management by direction and control, whether hard or soft, is inadequate to motivate because the human needs on which this approach relies are relatively unimportant motivators of behavior in our society today [McGregor, 1960, p. 42]." The self-control afforded participants in group decision making serves to satisfy higher level human needs.

Group decision making and other components of the human relations approach to organizational design rely heavily on Maslow's contribution. However, empirical evidence for the need hierarchy hypothesis is equivocal. While managerial research has demonstrated the importance of higher level needs and differential opportunities for their fulfillment (Porter, 1964; Porter & Lawler, 1968), field studies have supported only a two-step hierarchy of motives, with the social, ego, and self-actualization motives comprising the higher level (Hall & Nougaim, 1968; Lawler & Suttle, 1972).

Argyris' Individual-Organization Integration

While group decision making may be instrumental to the fulfillment of several motives, Argyris' (1957, 1964) principal focus is on its utility for self-actualization. The demands placed on people by the classical bureaucratic form of organization tend to thwart any movement toward self-actualization, which Argyris views as a complex motive comprised of: psychological success, self-awareness and control, independence, activity, personality growth, high aspiration levels, deep interests, and long-time perspectives. Where tasks are simplified, chains of command are pyramidal and rigid, spans of control are narrow, and so forth, the worker must behave in an environment where he has very little power, may develop a few shallow abilities and a short-time perspective, and is expected to be passive and dependent. Argyris recommends group decision making as one of several interdependent paths (including laboratory education in interpersonal competence, job enrichment, and reality-centered leadership) to the resolution of individual-organization conflict. When group decision making involves issues relevant to participants' work roles, it acts as a mechanism for increasing responsibility, autonomy, and achievement. Members can view the system in broader perspective, use their interaction and creative skills, and develop multidirectional power relationships.

Power Continuum and Expectations

Blake and Mouton (1961) have studied a "power spectrum" ranging from complete leader power (1/0) through collaboration (.5/.5) to complete subordinate power (0/1). Their research showed that as dyadic power relationships went from 1/0 to 0/1, (*a*) leaders' and subordinates' perceptions of responsibility changed curvilinearly with both maxima occurring at .5/.5; (*b*) subordinates' satisfaction increased linearly, and leaders' satisfaction decreased linearly. The optimum point on the spectrum, in terms of maximizing the felt responsibility of both leaders and subordinates and of equating their satisfaction, was a condition of power equalization.

The power expectation hypothesis suggests that affective responses should be a function of differences between power expectations and the extent to which these expectations are met. Horowitz (1964) varied expected and actual power relationships and examined the effects of four comparisons on subordinates' affect: (*a*) .5/.5 superior expectation, 1/0 subordinate expectation, and .5/.5 actual distribution; (*b*) .5/.5, .5/.5, and .5/.5; (*c*) 1/0, 1/0, and 1/0; and (*d*) 1/0, .5/.5, and 1/0. Liking for the superior was greatest when 1/0 expectations were met, second highest when subordinates expected no power and exercised equal power, and lowest when they expected equal power and exercised none. The optimal discrepancy between expected and actual power remains an empirical question. Likert (1961) suggests that this discrepancy should be slightly positive. That is, the subordinate should exercise more power than he expected, but not so much as to exceed his decision-making capabilities.

Likert's Participative Group Management

Likert (1961, 1967) considers group decision making an essential feature of management. His early research indicated that effective organizational units, as contrasted with ineffective units, (*a*) had employee-centered rather than job-centered leaders; (*b*) maintained general rather than close supervision; and (*c*) allowed freedom for the work group to set its own goals and pace. Satisfied workers valued group decision making and reported that their supervisors asked for their ideas and frequently acted on them. Power sharing is part of Likert's leadership principle of "supportive relationships," which underlies effective communication and motive fulfillment. He has presented data from business, industrial, voluntary, and government organizations which support the efficacy of his "System 4" or participative group form of organization (Likert, 1967).

Tannenbaum's Influence Pie

Tannenbaum (1968) has presented one of the more provocative theoretical innovations concerning power sharing. He argues that the total amount of power in an organization need not be conceptualized as a fixed quantity, where changes in power relationships necessitate a give-and-take process.

The fixed pie assumption has prevailed in previous research. The expectation model considers the consequences of varying *relative proportions* of a fixed power pie allocated to two persons or organizational levels. Blake and Mouton's (1961) power equalization, or

.5/.5 distribution, involves each person exercising *half* of the total power. Mechanic (1962) noted that decision strategies involve value judgments on the merits of retaining *or* delegating power. Tannenbaum, however, suggests that total group or organizational power is variable and that proper management practices can expand the pie.

Nonsupervisory personnel in 31 departments of a service organization indicated on independent rating scales the influence that they, their supervisors, and department managers exercised over department activities. Respondents attributed greater influence to successively higher levels. However, for the top one-third of the departments in terms of productivity, as contrasted with the lower third, influence was higher at all three organizational levels (Likert, 1961, pp. 55–56). Support has also been found for the hypothesis of expanded influence pies in Leagues of Women Voters (Tannenbaum, 1961), labor unions (Tannenbaum & Kahn, 1958), insurance agencies (Bowers, 1964), utility plants (Mann & Hoffman, 1960), and a factory in Yugoslavia (Zupanov & Tannenbaum, 1968).

Mechanisms of influence pie expansion are related to Likert's interaction-influence system. When a leader shares power, his own power may increase through his interaction with followers and acceptance of their contributions, which build a base for subordinates' responsiveness to the leaders' structuring behaviors. If the leader maintains influence over his superiors (Pelz, 1952), his group can increase its total power. This theoretical position also assumes that subordinates' influence in decision making is instrumental to decision implementation at their organizational level. Clarification of these mechanisms requires additional data.

CRITICISMS OF POWER SHARING

Although many contemporary organization theories endorse the sharing of social power, this thesis is not universally accepted. Several contemporary writers have tried to refute or modify the power-sharing thesis. Gilman (1962), arguing that power sharing fails to limit workers' power-seeking behavior, advocates a middle-of-the-road approach where power is exercised through both formal and informal organization processes and is granted from below as well as delegated from above. Leavitt (1965) concluded that the power-sharing propositions seem essentially correct, but they have been overgeneralized and lack empirical support. Much of the power-sharing research is inadequate either in its specification of power-sharing techniques or its evaluation criteria. Sensitivity training studies elaborate intragroup processes but rarely invoke adequate criteria (Campbell & Dunnette, 1968). Neither early field experiments on group decision making (Coch & French, 1948; Morse & Reimer, 1956) nor research conducted in a correlational paradigm (Likert, 1961) has isolated power sharing per se as a cause of organizational effectiveness (Wood, 1970, Ch. 1).

Some concomitants of power sharing may also be incompatible with its goals. Power sharing may generate cohesiveness and loyalty and inhibit creativity. The unlimited communication of information across levels may exceed the information-filtering and selectivity processes in the organization. Unrestrained power sharing may also ignore individual differences in the motivation to exercise power.

Strauss (1963) has presented the most comprehensive critique. His basic points can be summarized as follows: (*a*) The assumptions about human motivation on which power sharing rests are inaccurate. Such assumptions are based on value judgments made in academia and rarely converge with workers' values. In particular, economic motivation is underemphasized. Recent discussions of the instrumentalities of economic reward for higher level need fulfillment weaken this criticism (Lawler, 1971; Opsahl & Dunnette, 1966). However, Strauss suggests that we underplay job satisfaction and accelerate automation to provide more leisure time in which man can strive to self-actualize. (*b*) Power sharing ignores organizational costs. Gains attributable to power sharing may not compensate for the costs of its implementation. (*c*) Power-sharing programs produce dysfunctional consequences, specifically: (*i*) Resulting cohesiveness may work against organizational goals; (*ii*) an initial exercise of power may generate expectations for continuous participation, which management may not be able to meet; (*iii*) participants who do not exercise power, for example, who have their ideas rejected, may become hostile and alienated; (*iv*) group decision making can be frustrating and generate decisions of questionable quality; (*v*) group decision making can readily become a manipulative tool of management to attain its ends.

Many of these criticisms seem valid. Yet, it would be as premature to totally subscribe to them as it would be to adopt power sharing as an organizational panacea. Resolution of theoretical conflict concerning power relationships and group decision making requires research on the nature and consequences of group decision-making processes and stipulation of boundary conditions for power sharing.

SITUATIONAL-PROCESS APPROACH

Group Decision-Making Process

Group decision making has been the most extensively studied means of sharing power. However, neither the operational referents of this process nor the components responsible for its effects are clearly defined. The approach diagrammed in Figure 1 suggests consideration of three elements of the process: (*a*) participation in phases of decision making, (*b*) multiple power inputs, and (*c*) interpersonal interaction in the process.

Participation in decision process phases. Researchers have defined participation as: joint decision making by two or more parties; decision making by subordinates; the amount of talking or other demonstrative behaviors; and the amount of influence exercised.[2] The assumption of participation–influence isomorphism has been particularly prevalent in research on decision-making styles. Tannenbaum and Schmidt (1958) discussed seven styles, ranging from the leader's making a decision and announcing it, to his permitting subordinates to contribute within leader-defined limits. These styles varied (conceptually) along a leader–power continuum. Heller and Yukl (1969) also placed delegative, consultative, and joint decision-making styles along an influence continuum. Such assumptions will be valid, however, only to the extent that manipulations of participation relate to the exercise of influence and to members' perceptions of influence relationships.

An alternative approach conceptualizes group decision making as a process comprised of phases in which organizational members may have an opportunity to take part (participate). Based on theories of executive decision making (Bass, 1965; Cyert & March, 1963; March & Simon, 1958) and on empirical assessments of problem-solving interactions (Bales & Strodtbeck, 1951), a tentative formulation of the process incorporates three phases—generation, evaluation, and choice of alternative resolutions of a decision issue. Variations of group decision making involve complete-process participation and forms of group decision making where subordinates take part in only one or two phases. Partial group decision making is common in organizational practice. A leader may submit a fixed set of alternatives to group evaluation; he may seek generation of alternatives through "suggestion box" programs; or he may involve subordinates in only generation and evaluation, as implied in the following statement:

> I would not think of making a decision by going around the table and then deciding on the basis of how everyone felt. Of course, I like to hear everyone, but then I go off alone and decide. The decisions that are important must be made alone [R. M. Nixon, as quoted by Schecter, 1972, pp. 18–19].

Vroom (1973) has incorporated a similar decision-phase concept in developing a model of organizational decision making. He presented five decision styles for group issues, ranging from "You solve problem or make decision yourself using information available . . ." to "You share problem with your subordinates as a group. Together you *generate* and *evaluate* alternatives and *attempt to reach agreement* (consensus) on a solution [italics added]." Intermediate styles involve the leader's gathering information from subordinates individually, or allowing subordinates, individually or as a group, to generate (and perhaps evaluate) alternatives.

---

[2] J. Anderson. The effect of participation and consideration on subordinate job attitudes. Unpublished manuscript, University of Illinois, 1969.

With this framework, one can examine the effects of subordinates' participation in decision phases on the development of power relationships. Laboratory research has shown that complete-process group decision making generates higher levels of subordinate influence than does partial group decision making. This effect also holds for the group influence pie; subordinates in complete and partial group decision making conditions reported similar levels of leader influence (Wood, 1970, 1972a). Influence also depends on the particular phase in which members participate. Group members who participated only in generating alternatives perceived their influence as slightly higher than did members who only evaluated alternatives; members involved only in final alternative choice described their influence as significantly greater than did participants in either of the other phases.[3]

Multiple power inputs to group decision making. Participants in group decision making may exercise influence derived from multiple bases of power. Social power, or one's capability of influencing (i.e., affecting the behavior of) another (Cartwright, 1965), develops from several sources. These multiple bases provide different rationales for one's compliance with the influence attempts of another. Weber (1961) argued that power vested in bureaucratic positions was of three "types" —legal authority legitimized by rational law and its formal application, traditional authority derived from the sacredness of the social order, and charismatic authority based on followers' devotion to leaders' heroic qualities.

Barnard (1938) introduced the distinction between "authority of position" and "authority of leadership," which Bass (1960) rephrased in terms of "position power" and "personal power." Interpersonal exchange theory distinguishes fate power and behavior power (Thibaut & Kelley, 1959). Fate power refers to the automatic effect of one's behavior on the interaction outcome for another person; in exercising behavior power, one varies his behavior to make it desirable for the other person to change his behavior. Etzioni (1965) has suggested classifying organizations according to their dominant power base. The type of organizational commitment will depend upon whether the primary power is (*a*) coercive, based on physical sanctions; (*b*) utilitarian, based on material rewards; or (*c*) identitive, based on identification with reference group norms.

French and Raven's (1959) description of five power bases is probably the most widely accepted version of multiple power bases. Legitimate power derives from authority assigned to positions. Legitimate power may be greater in positions of centrality (Mechanic, 1962). Power may be based on expertise gained through experience and training. A leader may exercise referent power if followers consider him a member of their primary reference group. Hollander and Webb (1955) and Hurwitz, Zander, and Hymovitch (1968) have substantiated an association between interpersonal attraction and the exercise of power. Referent power is more likely, however, to be a source of subordinate influence exercised through coalitions or the informal organization (Mechanic, 1962). Coercive power, based on the application of physical force, is another possible source of power, though it rarely occurs in modern organizations. Control of rewards provides a common power base through which workers' needs may be satisfied. This source of power may derive from task-environmental rewards or rewards associated with successful interaction (Collins & Guetzkow, 1964). These multiple power bases may interact in determining one's influence over another person or the products of group decision making.

Multiple power bases relate differently to organizational consequences. To the extent that managers were reported to rely on referent and expert power bases, their subordinates' satisfaction and performance were high. Satisfaction was inversely related to reward, coercive, and legitimate power. Reward power also correlated negatively with performance (Bachman, Smith, & Slesinger, 1966). Studies of salesmen, college faculties, life insurance agents, and production workers replicated the relationships of referent and expert power to satisfaction (Bachman, Bowers, & Marcus, 1968). However, reward power correlated positively with satisfaction and performance in insurance agencies. Among production workers, satisfaction was directly related to legitimate power, but performance was positively related to referent power.

---

[3] M. R. Cooper. Satisfaction, perceived influence, and decision-riskiness: The effects of participation and commitment in decision-making groups. Unpublished doctoral dissertation, Ohio State University, 1972.

Further work is needed on the effects of power bases and their development. Effects of group decision making on power sharing and organizational effectiveness may depend on the interaction of multiple power bases across individuals and hierarchical group levels. Research could be designed to test the hypothesis that the effects of power bases should interact with phases of the group decision-making process. It appears reasonable to predict that expertise would be beneficial to generating high-quality decision alternatives, that successful group evaluation would rely more heavily on referent power, and that choice behavior would depend on reward and legitimate power.

Interaction in the group process. The present model also considers the extensiveness and qualitative nature of interpersonal interaction as important determinants of group decision-making effects. Distinctions between decision styles imply different degrees of leader–subordinate interaction. Individual leader decision making and delegation should yield different power relationships though neither approach involves extensive interaction. Interaction in group decision making may, however, affect differences in intersubordinate, as well as leader–subordinate power relations. Interaction should also be critical in the development of the additional intervening processes of interpersonal affiliation and trust, which in turn should relate to organizational consequences.

Research on choice shifts in decision behavior indicates that the degree of interaction also affects the magnitude of choice shifts. Group decision making generally produces decisions of greater risk than does individual decision making. Several interaction-based mechanisms, such as diffusion of responsibility and social comparison of values, have been offered as explanations for this phenomenon (Pruitt, 1971).

Evidence also suggests that the utility of interpersonal interaction depends on its application to different phases of the decision process. Vroom, Grant, and Cotton (1969) found interaction to be functional in evaluating decision alternatives, but not in generating them. Qualitative interaction differences should also affect power relationships, since group members perceive different behaviors exhibited in group decision making to reflect different degrees of influence (Wood, 1970).

Situational and Individual Differences

The second basic focus of the situational-process approach is on the role of situational and individual variables. These classes of variables should moderate both the effects of group decision making on intervening processes, such as power and affiliative relationships, and the relationships between these processes and organizational outcomes, such as attitudes and performance.

Group decision-making theorists have recognized the general role of situational constraints. McGregor (1960) introduced the concept of selective adaptation wherein appropriate control mechanisms are adapted to the situation.[4] He suggested that group decision-making effectiveness will depend on managers' skills, the organizational level where decision making occurs, and the nature of the decision issues (Ch. 9). Maier (1955) has similarly indicated that the success of group decision making will be bounded by an "area of freedom," defined in part by the nature of the task and the discussion skills of the group leader. Argyris noted individual differences in motivation to exercise power (1957) and, more recently, has defined some organizational conditions requiring different decision-making structures and processes (1971). Blake and Mouton (1961) advocate their collaborative power relationship when interaction opportunities exist. Subordinates should exercise greater power if they are exceptionally skilled; greater leader power is necessary in times of crisis or when subordinates lack skills and information.

---

[4] Situational considerations were evident much earlier in the work of Follett (Metcalf & Urwick, 1942), who attempted to incorporate employee representation and decentralization in classical structural theory. Her "law of the situation" implied that leaders must diagnose situations and elicit behavior compatible with their demands in order to effectively exercise power.

Theory and research should identify and integrate critical mediators, such as those variables associated with the organizational environment, decision tasks, and motives of participants.

Organizational characteristics. There has been little research on effects of specific situational variables in the organizational context on power relationships. However, research and theory on broad organizational parameters suggest that characteristics such as the general mission of the organization (Sloan, 1964), interdependence of functional subsystems (Marschak, 1959; Porter & Lawler, 1965), technology and programming of tasks (Woodward, 1965), and managerial climate (McGregor, 1960) may affect the utility of power sharing. Through their effects on organizational structures and processes, these parameters may generate extant balances of power that preclude power sharing. Mulder and Wilke (1970) assume that group decision making will not lead to power sharing, given an expert power relationship favoring the leader. Rather, the leader–subordinate power gap will widen unless the subordinates develop expertise by interacting with an expert leader. Mulder and Wilke argue that effective group decision making requires participant equality on abilities and motivational dimensions. The exercise of power in decision making may also enhance expectations of exercising power in the future and thereby may limit organizational flexibility in decision-making approaches (Mulder, 1959). Shiflett and Nealey (1972) manipulated the legitimate power of the leaders of task groups. When his power was augmented, the leader engaged in more structuring, directive behaviors. However, when the power manipulation was reversed, power-related behaviors did not decrease.

Several identifiable organizational characteristics should limit the effectiveness of group decision making. Theoretically, group decision making should be ineffective to the extent that (*a*) few organizational levels are subsumed by the group decision-making process; (*b*) little useful information is available at lower levels; and (*c*) time pressures on decision making are severe (Lowin, 1968). Leavitt (1965) has also noted that organizational size and complexity may impose constraints on the feasibility of group decision making.

Heller and Yukl (1969) surveyed the extent to which managers operating under different organizational conditions used the following decision-making approaches: (*a*) own decision without explanation to subordinates, (*b*) own decision with explanation, (*c*) consultation, (*d*) joint decision making, (*e*) delegation. Decision styles varied by organizational level and function. Student leaders most frequently used *d* above, senior business managers emphasized *e*, second-level supervisors used *c* and *b*, and *b* was the most common method for first-line supervisors. Personnel and general managers relied more on group decision making than did production and finance managers. Greater spans of control were associated with Method *b* for second-level supervisors, and with either *a* or *e* in top management. Job experience was related to *c* at upper levels and *a* at lower levels. Lower level managers more often applied group decision making to task than to group maintenance types of decisions. These data, while addressed to group decision-making utilization rather than outcomes, indicate a major role for situational moderating variables.

Decision task. A more general underlying situational concept may be associated with decision tasks. Wood (1972a) distinguished decision tasks as being facilitative or contrastive, based on Deutsch's (1949) consideration of promotively interdependent and contriently interdependent goal structures. Deutsch's promotive case implied cooperation in that an individual cannot attain his goal unless other group members attain theirs. A contrient goal structure, where the individual can reach his goal only if other members fail to reach theirs, generates intragroup competition. Wood's facilitation–contrast was less stringent than the promotion–contrience definition. Contrast was based on differing member preferences for decision methods to be applied to particular issues. Contrast could be extended to incorporate differing preferences for the nature of the resulting decision. Facilitation referred to the case where, by virtue of compatible reference-group norms for shared power in resolving the issue, or convergence of outcome preferences, the task facilitated the mutual exercise of power by superiors and subordinates. These formulations imply that facilitation–contrast depends both on the specific issue under consideration and the decision-making participants. A contrastive situation is similar to Vidmar's (1967) conceptualization of a negotiation group, where members belong to reference groups with conflicting views, and these role forces compete with the force toward generating acceptable, constructive decisions. Lowin (1968) has suggested that expansion of the influence pie is more probable when leaders and subordinates are not contriently related.

Wood's data supported the hypothesis that power sharing should be more evident with facilitative than with contrastive tasks. Subordinates in male decision groups reported greater subordinate and leader influence in facilitative task situations. Subordinates' assessments also indicated that "power equalization" (no significant difference between leader and subordinate influence) occurred only when subordinates participated in the complete (generation-evaluation-choice) decision process of resolving facilitative task issues. Further work on the determinants of facilitation should examine task properties generating facilitative relationships, individual characteristics associated with goal structures, and attitudes toward decision-making roles.

The work of Vroom (1973) should increase our understanding of the situational facilitation of group decision-making approaches. He has examined the use of variations of decision-making strategies in situations involving different "problem attributes." One attribute refers to the likelihood of conflict among subordinates in preferred solutions. This corresponds closely to Wood's preliminary definition of task facilitation–contrast. In addition, however, Vroom has incorporated in his model other potentially critical attributes of the problem, including intermember trust to base solutions on organizational considerations, the importance and likelihood of acceptance and decision implementation, the importance of decision quality, and the importance and accessibility of information held by various members. Normative rules specify the appropriateness of decision styles given the problem conditions. These rules generate a set of feasible approaches that protect both quality and acceptance criteria. One can then select an approach from the set of possible styles on the basis of minimizing man-hour investments in the decision process. Data from a cross-cultural sample of managers indicated that decision makers frequently violate the proposed decision rules, particularly those protecting acceptance. The further results of Vroom's work, including tests of the fit between his normative model and descriptive models, and the training of managers to diagnose problem attributes, will deserve the attention of group decision-making theorists.

The foregoing research implies that for participative management to be effective, either the group approach must be limited to subsets of situations, or long-term involvement in such a system should reduce the contrast of decision tasks through the development of mutual trust and openness. Evidence from a field experiment with 10–14-year-old boys in athletic training showed greater effects of participation in later trials (Veen, 1972). However, intervening processes operating over time were not cleanly isolated. Longitudinal studies are especially critical in light of recent evidence that group power structures interact with task difficulty and the time of criterion measurement in their effects on group performance (Levine & Weitz, 1971). Groups with both balanced control structures and high levels of total group control performed best in the later stages of *hard* problem-solving tasks. However, while balanced control groups were also superior to unbalanced groups in the late part of *easy* tasks, the differentiation of amounts of control had its effect in the early trials, where high-control groups outperformed low-control groups. Moreover, in addition to Lowin's (1968) prescription that group decision tasks be directly relevant to the participants, the work of Levine and Weitz indicates that differences in power distribution become more important on difficult tasks.

Individual motives. Another class of variables that may account for variance in the effectiveness of group decision making involves participants' motivational differences. Individual differences in motivation seem most directly relevant to attitudinal effects, which will depend on motive structures and the opportunity provided for motive attainment by group decision making. Previous research supports this general proposition. Vroom (1960) measured individual needs for independence and authoritarianism, perception of participation, overall job satisfaction, and job performance among managers in a national delivery company. Significant correlations between participation and attitudes were obtained only for subsets of the sample with strong and moderate independence needs, and moderate and weak authoritarianism needs. Similar relationships occurred on the performance criterion, although between-subset correlation differences were not significant.

Differences attributed to specific motives may also be reflected in broader cultural values. French, Israel, and Ås (1960) attempted to replicate the Harwood experiment on resistance to change (Coch & French, 1948) in a Norwegian factory. The expected beneficial results of group decision making did not occur, presumably, in part, because of the incompatibility of cultural norms with lower-level workers' acceptance of decision-making responsibility. Hulin and Blood (1968) have introduced the concept of "alienation from middle-class norms" to explain urban, blue-collar workers' negative reactions to efforts to enrich their jobs. Workers who are alienated from middle-class, Protestant-ethic work values, as indexed by such community variates as standard of living and population density, tend not to be satisfied with job enrichment programs. Inasmuch as such programs involve extending responsibility to subordinates, these generalizations may apply to other forms of power sharing, such as group decision making. Etzioni (1965) has suggested this extension, in stating: "It is in factories where alienation is not high to begin with . . . that organizational efforts to increase commitment and to control expressive activities through such mechanisms as . . . participation . . . in decision making are successful [pp. 668–669]."

While these approaches suggest that we can determine subsets of the population which will respond favorably to power sharing, it is also possible that power-sharing programs will generate attitude or behavioral change within a heterogeneous group based on multiple motivational mechanisms. The strength of participants' power motivation or related ego motives should predict responses to power sharing. Lowin (1968) notes that economic motives can also be met if participation is associated with interdependence of financial rewards. One might therefore expect favorable responses to group incentive systems, particularly those which are participatively developed. Yet, when group participation and individual reward are combined in a system, it is of interest to examine the contributions of each component to group effectiveness. Recent work underlines the importance of the power-sharing element. Among building-maintenance workers, a participatively developed incentive system significantly improved job attendance rates, while the same system (equivalent amount of and basis for reward) imposed by management failed to elicit behavior changes (Lawler & Hackman, 1969; Scheflen, Lawler, & Hackman, 1971).

Group processes designed principally to redistribute power may also satisfy social motives. Affiliation can accrue from group decision making, particularly when members interact over time and under conditions which facilitate agreement and joint goal-seeking activities. Wood[5] (1970, 1972b) has shown that both power and affiliation motives moderate attitude outcomes of group decision making. Influence perceptions were related to satisfaction only for group members with strong and moderate power motives. For individuals with strong affiliation motives, perceived participation (involvement in the group process) was significantly related to satisfaction, even though these attitude measures were not related to the exercise of influence. The mediating effects of social motives also interact with the decision process and task parameters previously discussed. Facilitative decision tasks were more conducive to satisfaction than contrastive tasks for people with strong affiliation motives. Moreover, participation in the choice phase of the decision process was related to satisfaction for group members with weak, but not those with strong, affiliation motives (Wood, 1970). Characteristics of situations, decision processes, and the people involved in the process must be considered as interacting determinants of power distribution and its effects.

---

[5] M. T. Wood. Some determinants and consequences of power distribution in decision-making groups. Paper presented at the meeting of the American Psychological Association, Washington, D.C., September, 1971.

SUMMARY AND GUIDELINES FOR FUTURE RESEARCH

This review of power sharing and group decision making implies a need to conceptualize organizational power relationships in terms of a model which will account for the psychological mechanisms underlying shifts in power distributions and which defines the boundary conditions for the useful implementation of power-sharing programs in organizational development. A basic implication of the preliminary situational-process model is that a viable theory of power in organizations must deal with differences in situations and individuals who perform organizational roles. This approach is consistent with previous models of leadership processes. Characteristics of organizational environments, decision tasks, and individual motives will affect both perceived distributions of power resulting from organizational change programs and individuals' affective responses to power sharing. Further research is needed to isolate the specific variables within these broad categories of boundary conditions which contribute to variance in power distribution and behavior, to systematically test their interactions, and to develop basic theoretical dimensions of environments, tasks, and people.

A related implication involves the conceptualization of power as an intervening process in organizational development, rather than as a structural given or terminal effect of planned change. Power relationships are viewed as deriving from multiple inputs, in terms of decision-making strategies and other organizational design variables, and as leading to multiple outputs in terms of organizational consequences such as member attitudes and behaviors. The situational elements of the model may differ in their effects of mediating input-power and power-output relationships. Structural and task mediators may be most important in the former linkage, and motivational mediators in the latter. Further, specific motivational mediators may affect both linkages. Power motivation affects individual influence attempts,[6] as well as attitudes resulting from the successful exercise of influence. The model also incorporates other intervening processes associated with group decision making. Affiliative processes may relate to social motive fulfillment (Wood, 1972b). Interpersonal trust should enhance commitment and decision implementation (Lawler & Hackman, 1969).

Examination of these intervening processes is crucial if we are to understand the dynamics of organizational change. Many change programs involving role design and personnel development are intended to directly affect power relationships. On the other hand, programs may be designed to improve role attitudes and behaviors but fail to pay explicit attention to intended or serendipitous changes in intervening processes. Job enrichment and human relations training are often adopted as treatment packages to generate change. However, the focus is generally on the relation between the treatment and the organizational consequence rather than on intervening mechanisms which might yield changes in the consequence. Moreover, programs designed to affect organizational behavior *through* changes in power relationships may generate changes in behavioral outcomes, but *not* via the power mechanism. In a reported 2-year program incorporating group decision making, earnings development, and training, performance improved 30%, but there were no significant changes in power relationships (Marrow, Bowers, & Seashore, 1967).

The present approach also suggests examination of effects of power relationships on multiple organizational consequences. Criteria for group decision-making effects have varied, and several studies have shown different treatment effects on different criteria (Marrow et al., 1967; Morse & Reimer, 1956). It may be helpful to distinguish classes of consequences based on immediacy of outcomes. The more proximal class includes characteristics of decisions, particularly their quality, riskiness, and acceptance by organizational members. Group decision making purportedly produces an optimal balance between decision quality and acceptance. Maier (1955) has noted that acceptance is the major benefit of group decision making. Further, to the extent that participants demonstrate task expertise, decisions should be of high quality. Reported data support the acceptance hypothesis in terms of compliance and the execution of decisions. The quality of group decisions, however, has not been clearly documented, and is difficult to evaluate without quality assessments independent of the group participants.

---

[6] D. Courtney. The effects of member power, power motivation, and set to cooperate or compete with other group members in small decision-making groups. Unpublished doctoral dissertation, Ohio State University, 1971.

The more distal set of consequences of group decision making and power sharing includes organizational role performance and attitudes, and other measures of system effectiveness. Predictions of attitudes follow from psychological theory underlying power sharing if we formulate hypotheses based on motive fulfillment as an attitude determinant and consider the multiple motivational mechanisms operating in group decision-making strategies. Relationships of group decision making and power sharing to role performance should depend on additional organizational parameters, such as decision–reward contingencies and individual expectancies (Lawler, 1971; Vroom, 1964, 1970). Contributions of power-sharing approaches to system effectiveness should depend on the longevity and pervasiveness of the approaches and the feedback of their consequences to future decision procedures and organizational development programs.

Future research efforts can be directed toward the conceptual development of the components of the situational-process model, their effective manipulation or measurement, and their causal linkages. Effective organizational design rests upon an understanding of individual and situational mediators of effects of decision processes and power relationships. The validity and complexity of the model will influence the utility of situational change or adaptive treatment strategies. Adapting decision methods to fit individual–organizational contexts assumes flexibility in organizational roles and effective means of assessing the context variables. On the other hand, the consistent use of power-sharing methods may require extensive change in organizational structure, decision problems, and individual attitudes. The resolution of the theoretical and practical issues concerning organizational power relationships will come only from systematic laboratory research designed to identify and link causal relationships, and from organizational applications designed to discover critical interactive variables and to evaluate the comparative utility of different decision-making approaches and power distributions.

REFERENCES

Argyris, C. *Personality and organization.* New York: Harper, 1957.

Argyris, C. *Integrating the individual and the organization.* New York: Wiley, 1964.

Argyris, C. *Management and organizational development: The path from XA to YB.* New York: McGraw-Hill, 1971.

Bachman, J. G., Bowers, D. G., & Marcus, P. M. Bases of supervisory power: A comparative study in five organizational settings. In A. S. Tannenbaum (Ed.), *Control in organizations.* New York: McGraw-Hill, 1968.

Bachman, J. G., Smith, C. G., & Slesinger, J. A. Control, performance, and satisfaction: An analysis of structural and individual effects. *Journal of Personality and Social Psychology,* 1966, **4**, 127–136.

Bales, R. F., & Strodtbeck, F. L. Phases in group problem-solving. *Journal of Abnormal and Social Psychology,* 1951, **46**, 485–95.

Barnard, C. I. *The functions of the executive.* Cambridge: Harvard University Press, 1938.

Bass, B. M. *Leadership, psychology, and organizational behavior.* New York: Harper, 1960.

Bass, B. M. *Organizational psychology.* Boston: Allyn & Bacon, 1965.

Bennis, W. G. A funny thing happened on the way to the future. *American Psychologist,* 1970, **25**, 595–608.

Blake, R. R., & Mouton, J. S. *Group dynamics—key to decision making.* Houston: Gulf, 1961.

Bowers, D. G. Organizational control in an insurance company. *Sociometry,* 1964, **27**, 230–244.

Campbell, J. P., & Dunnette, M. D. Effectiveness of T-group experiences in managerial training and development. *Psychological Bulletin,* 1968, **70**, 73–104.

Campbell, J. P., Dunnette, M. D., Lawler, E. E., & Weick, K. E. *Managerial behavior, performance, and effectiveness.* New York: McGraw-Hill, 1970.

Cartwright, D. Influence, leadership, control. In J. G. March (Ed.), *Handbook of organizations.* Chicago: Rand McNally, 1965.

Coch, L., & French, J. R. P., Jr. Overcoming resistance to change. *Human Relations,* 1948, **1**, 512–533.

Collins, B. E., & Guetzkow, H. *A social psychology of group processes for decision-making.* New York: Wiley, 1964.

Cyert, R. M., & March, J. G. *A behavioral theory of the firm.* Englewood Cliffs: Prentice-Hall, 1963.

Deutsch, M. A theory of cooperation and competition. *Human Relations,* 1949, **2**, 129–152.

Dunnette, M. D. *Personnel selection and placement.* Belmont, Calif.: Wadsworth, 1966.

Etzioni, A. Organizational control structure. In J. G. March (Ed.), *Handbook of organizations.* Chicago: Rand McNally, 1965.

Fayol, H. *General and industrial management.* London: Pitman & Sons, 1949.

Fiedler, F. E. *A theory of leadership effectiveness.* New York: McGraw-Hill, 1967.

French, J. R. P., Jr., & Raven, B. The bases of social power. In D. Cartwright (Ed.), *Studies in social power.* Ann Arbor: University of Michigan, Institute for Social Research, 1959.

French, J. R. P., Jr., Israel, J., & Ås, D. An experiment on participation in a Norwegian factory. *Human Relations,* 1960, **13**, 3–19.

Gilman, G. An inquiry into the nature and use of authority. In M. Haire (Ed.), *Organization theory in industrial practice.* New York: Wiley, 1962.

Gouldner, A. W. *Patterns of industrial bureaucracy.* Glencoe, Ill.: Free Press, 1954.

Gulick, L. Notes on the theory of organization. In L. Gulick & L. Urwick (Eds.), *Papers on the science of administration.* New York: Columbia University, Institute of Public Administration, 1937.

Hall, D. T., & Nougaim, K. E. An examination of Maslow's need hierarchy in an organizational setting. *Organizational Behavior and Human Performance,* 1968, **3**, 12–35.

Heller, F. A., & Yukl, G. Participation, managerial decision making, and situational variables. *Organizational Behavior and Human Performance,* 1969, **4**, 227–241.

Hollander, E. P., & Julian, J. W. Contemporary trends in the analysis of leadership processes. *Psychological Bulletin,* 1969, **71**, 387–397.

Hollander, E. P., & Webb, W. B. Leadership, followership, and friendship: An analysis of peer nominations. *Journal of Abnormal and Social Psychology,* 1955, **50**, 163–167.

Horowitz, M. Conflict in organization. In R. L. Kahn & E. Boulding (Eds.), *Power and conflict in organizations.* New York: Basic Books, 1964.

Hulin, C. L., & Blood, M. R. Job enlargement, individual differences, and worker responses. *Psychological Bulletin,* 1968, **69**, 41–55.

Hurwitz, J. I., Zander, A. F., & Hymovitch, B. Some effects of power on the relations among group members. In D. Cartwright & A. F. Zander (Eds.), *Group dynamics: Research and theory.* New York: Harper & Row, 1968.

Katz, D., & Kahn, R. L. *The social psychology of organizations.* New York: Wiley, 1966.

Lawler, E. E., III. *Pay and organizational effectiveness: A psychological view.* New York: McGraw-Hill, 1971.

Lawler, E. E., III., & Hackman, J. R. Impact of employee participation in the development of pay incentive plans: A field experiment. *Journal of Applied Psychology,* 1969, **53**, 467–471.

Lawler, E. E., III, & Suttle, J. L. A causal correlational test of the need hierarchy concept. *Organizational Behavior and Human Performance,* 1972, **7**, 265–287.

Leavitt, H. J. Applied organizational change in industry: Structural, technological, and humanistic approaches. In J. G. March (Ed.), *Handbook of organizations.* Chicago: Rand McNally, 1965.

Levine, E. L., & Weitz, J. Relationship between task difficulty and the criterion: Should we measure early or late? *Journal of Applied Psychology,* 1971, **55**, 512–520.

Likert, R. *New patterns of management.* New York: McGraw-Hill, 1961.

Likert, R. *The human organization: Its management and value.* New York: McGraw-Hill, 1967.

Lowin, A. Participative decision making: A model, literature critique, and prescriptions for research. *Organizational Behavior and Human Performance,* 1968, **3**, 68–106.

Maier, N. R. F. *Psychology in industry.* Boston: Houghton Mifflin, 1955.

Mann, F. C., & Hoffman, L. R. *Automation and the worker: A study of social change in power plants.* New York: Holt, Rinehart & Winston, 1960.

March, J. G., & Simon, H. A. *Organizations.* New York: Wiley, 1958.

Marrow, A. J., Bowers, D. G., & Seashore, S. E. *Management by participation.* New York: Harper & Row, 1967.

Marschak, J. Efficient and viable organizational forms. In M. Haire (Ed.), *Modern organization theory.* New York: Wiley, 1959.

Maslow, A. *Motivation and personality.* New York: Harper, 1954.

McGregor, D. M. *The human side of enterprise.* New York: McGraw-Hill, 1960.

Mechanic, D. Sources of power of lower participants in complex organizations. *Administrative Science Quarterly,* 1962, **7**, 349–364.

Merton, R. J. Bureaucratic structure and personality. In J. Litterer (Ed.), *Organization: Structure and behavior.* New York: Wiley, 1963.

Metcalf, H. C., & Urwick, L. (Eds.), *Dynamic* *administration: The collected papers of Mary Parker Follett.* New York: Harper, 1942.

Morse, N. C., & Reimer, E. The experimental change of a major organizational variable. *Journal of Abnormal and Social Psychology,* 1956, **52**, 120–129.

Mulder, M. Power and satisfaction in task-oriented groups. *Acta Psychologica,* 1959, **16**, 178–225.

Mulder, M., & Wilke, H. Participation and power equalization. *Organizational Behavior and Human Performance,* 1970, **5**, 430–448.

Opsahl, R. L., & Dunnette, M. D. The role of financial compensation in industrial motivation. *Psychological Bulletin,* 1966, **66**, 94–118.

Pelz, D. C. Influence: A key to effective leadership in the first-line supervisor. *Personnel,* 1952, **9**, 3–11.

Porter, L. W. *Organizational patterns of managerial job attitudes.* New York: American Foundation for Management Research, 1964.

Porter, L. W., & Lawler, E. E., III. Properties of organization structure in relation to job attitudes and job behavior. *Psychological Bulletin,* 1965, **64**, 345–355.

Porter, L. W., & Lawler, E. E., III. *Managerial attitudes and performance.* Homewood, Ill.: Irwin, 1968.

Pruitt, D. G. Choice shifts in group discussion: An introductory review. *Journal of Personality and Social Psychology,* 1971, **20**, 339–360.

Schecter, J. The private world of Richard Nixon. *Time,* 1972, **99**, Jan. 3, 18–19.

Scheflen, K. C., Lawler, E. E., III., & Hackman, J. R. Long-term impact of employee participation in the development of pay incentive plans: A field experiment revisited. *Journal of Applied Psychology,* 1971, **55**, 182–186.

Shiflett, S. C., & Nealey, S. M. The effects of changing leader power: A test of "situational engineering." *Organizational Behavior and Human Performance,* 1972, **7**, 371–382.

Sloan, A. P. *My years with General Motors.* New York: Doubleday, 1964.

Strauss, G. Some notes on power-equalization. In H. J. Leavitt (Ed.), *The social science of organization.* Englewood Cliffs: Prentice-Hall, 1963.

Tannenbaum, A. S. Control and effectiveness in a voluntary organization. *American Journal of Sociology,* 1961, **67**, 33–47.

Tannenbaum, A. S. (Ed.), *Control in organizations.* New York: McGraw-Hill, 1968.

Tannenbaum, A. S., & Kahn, R. L. *Participation in union locals.* Evanston, Ill.: Row, Peterson, 1958.

Tannenbaum, R., & Schmidt, W. H. How to choose a leadership pattern. *Harvard Business Review,* 1958, **36**, 95–101.

Taylor, F. W. *Scientific management.* New York: Harper, 1947.

Thibaut, J. W., & Kelley, H. H. *The social psychology of groups.* New York: Wiley, 1959.

Udy, S. H. Administrative rationality, social setting, and organizational development. In A. H. Rubenstein & C. Haberstroh (Eds.), *Some theories in organization.* Homewood, Ill.: Irwin, 1966.

Veen, P. Effects of participative decision making in field hockey training: A field experiment. *Organizational Behavior and Human Performance,* 1972, **7**, 288–307.

Vidmar, N. J. Leadership and role structure in negotiation and other decision-making groups. (Doctoral dissertation, University of Illinois). Ann Arbor, Mich.: University Microfilms, 1967. No. 68-1879.

Vroom, V. H. *Some personality determinants of the effects of participation.* Englewood Cliffs: Prentice-Hall, 1960.

Vroom, V. H. *Work and motivation.* New York: Wiley, 1964.

Vroom, V. H. Industrial social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology.* Vol. 5. (2nd ed.) Reading, Mass.: Addison-Wesley, 1970.

Vroom, V. H. Leadership. In M. Dunnette (Ed.), *Handbook of organizational psychology.* Chicago: Rand McNally, 1973, in press.

Vroom, V. H., Grant, L. D., & Cotton, T. S. The consequences of social interaction in group problem solving. *Organizational Behavior and Human Performance,* 1969, **4**, 77–96.

Weber, M. Bureaucracy. In H. H. Gerth & C. W. Mills (Eds.), *From Max Weber: Essays in sociology.* New York: Oxford University Press, 1960.

Weber, M. The three types of legitimate rule. In A. Etzioni (Ed.), *Complex organizations.* New York: Holt, Rinehart & Winston, 1961.

Wood, M. T. Some determinants and consequences of power distribution in decision-making groups. (Doctoral dissertation, University of Illinois) Ann Arbor, Mich.: University Microfilms, 1970. No. 71-15002.

Wood, M. T. Effects of decision processes and task situations on influence perceptions. *Organizational Behavior and Human Performance,* 1972, 7, 417–427. (a)

Wood, M. T. Participation, influence, and satisfaction in group decision making. *Journal of Vocational Behavior,* 1972, 2, 389–399. (b)

Woodward, J. *Industrial organization: Theory and practice.* London: Oxford University Press, 1965.

Zupanov, J., & Tannenbaum, A. S. The distribution of control in some Yugoslav industrial organizations as perceived by members. In A. S. Tannenbaum (Ed.), *Control in organizations.* New York: McGraw-Hill, 1968.

(Received March 30, 1972)

Psychological Bulletin
1973, Vol. 79, No. 5, 294–309

IMPLICIT PERSONALITY THEORY:
A REVIEW[1]

DAVID J. SCHNEIDER[2]

Brandeis University

A review of the literature on implicit personality theory reveals that there has been considerable concentration on various techniques for measuring the theory. While such measures do have considerable overlap they also vary in a variety of ways. The major theoretical questions have involved personality correlates of individual implicit personality theory, relationship of the theory to linguistic structure, and whether it reflects the actual distribution of traits in other people. Questions are raised about whether traits are the most appropriate units of person cognition and whether perceivers see traits as distributed across situations as well as stimulus persons.

In the first comprehensive review of research literature on person perception, Bruner and Tagiuri (1954) introduced the term "naive, implicit personality theory" to describe the possibility that perceivers assume inferential relationships among attributes of people. While the term was a new one, the idea was relatively old. Furthermore, quite similar ideas were developed about the same time by Jones (1954), Steiner (1954), Kelly (1955), and Cronbach (1955). It is the purpose of this paper to explore the varieties of ways the term has been used, to review the issues which have been raised in experimental studies, and to suggest a broadened notion of implicit personality theory.

Two traditions are represented in the notion of implicit personality theory. The first concerns the role of general bias in judgments of others, and the second has to do with individual differences in person perception. Various people who have examined trait ratings have noticed that perceivers tend to exaggerate the size of relationships among traits (Newcomb, 1931; Thorndike, 1920). This tradition has been represented on the one hand by those who see person perception and particularly impression formation as very much like concept formation or as similar to a wide variety of judgmental phenomena (e.g., Bruner, Shapiro, & Tagiuri, 1958; Wishner, 1960) and by those who have been primarily concerned with an analysis of linguistic factors in judgment (e.g., D'Andrade, 1965; Osgood, 1962). Thus in attempting to explain the pervasiveness of perceived trait covariation, researchers have looked to a presumably more basic psychology of human judgment or to biases implicit in our language.

A second major tradition has been more concerned with individual differences among perceivers. Research on authoritarianism, "the New Look," personality measurement, and personal construct theory all contribute to this interest.

The term "implicit personality theory" was initially used by Bruner and Tagiuri to describe assumed relationships of traits. Clearly, however, the term could be used to correspond to the broad conception of classic personality theory, namely, a set of assumptions about why people behave the way they do. Cronbach (1955) attempted to view implicit personality theory as more than relationships among traits; specifically he suggested that a person's "lay theory of personality" consisted of the mean and variance of each trait as well as the covariance among traits. Except for Gross (1961) there have been almost no explicit studies of the first two properties of trait ratings.

---

[1] Writing of this paper was facilitated by grants GS-1946 and GS-33242 from the National Science Foundation. A leave of absence provided by Amherst College was spent at Stanford University where many conversations with Albert Hastorf stimulated the writing of the paper. Also, Reuben Baron, Walter Mischel, and Terry Osborne were helpful in organizing a first draft, and David Hamilton and Seymour Rosenberg provided useful comments on a later draft.

[2] Requests for reprints should be sent to David Schneider, Department of Psychology, Brandeis University, Waltham, Massachusetts 02154.

Classic personality theories devoted considerable attention not only to the selection of basic terms at two or more levels of generality or abstraction, but also to determination of dynamic and causal relations among them. That does not, of course, mean that naive theories of personality differ in content and dynamic properties, but the possibility is worth attention. A later section of the paper will review the limited research on these aspects of implicit personality theory. The first sections, however, deal with research on the perceived relationships among traits.

MEASUREMENT OF TRAIT RELATIONSHIPS

General Procedures

While most implicit personality theory researchers have at the least obtained a matrix of trait relationships, there are considerable differences in how these relationships are generated and what further analyses are performed on this trait matrix. We may consider the following several points where differences arise: (*a*) assignment of traits to objects, (*b*) creation of the trait matrix, (*c*) interpretation of trait relationships, (*d*) reduction of the trait matrix to some more basic matrix, (*e*) interpretation of this basic matrix.

Assignment of traits. Some researchers begin with several perceivers assigning traits to stimulus objects. Most often, perceivers are asked to assign to each trait a magnitude for each stimulus object, for example, how much of a particular trait each stimulus object possesses. The stimulus objects are typically people and the final result is a Perceiver × Trait × Stimulus Person matrix.

Generally, the perceiver dimension is ignored either by considering the Trait × Object matrix separately for each perceiver, or by creating a composite Trait × Object matrix by summing or averaging across perceivers. Levy and Dugan (1960) have, however, created a Composite Trait × Perceiver matrix where each cell represents how many times each perceiver had assigned each trait to all stimulus objects.

The trait matrix. When researchers begin by having subjects assign traits to stimulus objects, correlations among traits may be calculated across stimulus objects or across perceivers from the Trait × Stimulus Person or Trait × Perceiver matrix. It is also possible to calculate profile distances among traits (Cronbach & Gleser, 1953) instead of correlations. In some instances this is more appropriate (Rosenberg & Olshan, 1970) since correlations are sensitive only to similarities of profiles between traits across stimulus objects and not to mean differences among traits.

Most researchers generate the trait relationship matrix directly. In some way, subjects are asked to assess rather directly the relationships among traits. There are a number of strategies for getting these data.

Some studies have relied on ratings of similarity (Messick & Kogan, 1966; Pedersen, 1965). The subject is presented with a pair of traits and asked to rate how similar or different the traits are.

Researchers have used likelihood of co-occurrence methods (e.g., Bruner, Shapiro, & Tagiuri, 1958; Hays, 1958; Lay & Jackson, 1969; Warr & Sims, 1965). Here the technique is, Given that a person has Trait X, how likely is it that he has Trait Y?

Hinkle[3] (also described in Bannister & Mair, 1968) obtained perceived causal judgments from subjects. Each subject was presented with a construct and asked if a person were to change in that construct (e.g., change from intelligent to stupid), on which others of the 19 constructs used would he also change. The result is a matrix of perceived cause–effect inference.

A fourth method used by Rosenberg (Rosenberg, Nelson, & Vivekananthan, 1968; Rosenberg & Olshan, 1970) obtains measures of co-occurrence. Each subject divides a total sample of traits into piles, each pile representing a distinct person. Each trait can be used only once although Rosenberg reports the findings are not altered greatly if traits are used several times. The co-occurrences are converted into disassociation scores which basically depend on the number of subjects who disagree on whether two traits go together. It should be pointed out that one could also study co-occurrences in various situations or roles or in any feature which serves to distribute traits (Rosenberg & Jones, 1972).

---

[3] D. N. Hinkle. The change of personal constructs from the viewpoint of a theory of implications. Unpublished doctoral dissertation, Ohio State University, 1965.

Reduction and interpretation of the trait matrix. Any of the correlational, distance, or strength-of-inference measures can be summarized through a wide variety of analytic techniques. Factor analysis is often used to reduce a matrix of trait intercorrelations. Any of the above measures which are interpreted as distances lead to a variety of multidimensional scaling analyses. Rosenberg and Sedlak (1972) has an excellent summary of the assumptions underlying these models.

Both factor analysis and multidimensional scaling lead to interpretations in terms of dimensions underlying the trait matrix. Ordinarily, of course, interpretation of these dimensions is performed by examination of traits which load at the extremes of the dimensions. More satisfactory analytic techniques have been worked out and used in recent years. Rosenberg et al. (1968) have advocated a regression technique in which some extraneous measure (such as a measure of social desirability) is used as a dependent measure and projections on the various dimensions become the independent measures. Interpretation of the multidimensional space then becomes possible through the pattern of multiple correlations with these dependent measures.

Distance data may be analyzed in ways which do not lead to dimensional summaries. For example, one reasonable alternative is a hierarchical organization. Cluster analysis (Johnson, 1967) has been used to provide such a summary of trait relationship data (Rosenberg & Jones, 1972). Interpretation of clusters can be done through analysis of variance; presumably those external traits which result in high within-trait homogeneity while distinguishing between clusters, that is, with high between-cluster relative to within-cluster variance, can be used to interpret clusters.

Individual Difference Measurement

Individual differences among perceivers can be studied using variants of most of the techniques discussed. One possibility is to divide perceivers into separate groups on some a priori criterion and to study whether the groups differ in trait relationships. Another method is to determine what individual differences are present in the data itself. The simplest technique would be to factor-analyze among perceivers (the correlations based on the extent to which the perceivers agree in how much of the various traits they see across all stimulus persons). There are also multidimensional scaling models which allow determination of individual viewpoints from the co-occurrence or similarity data themselves (cf. Rosenberg & Sedlak, 1972).

The more general point, however, is that when the researcher begins with a Perceiver × Stimulus Object × Trait data matrix, relationships among entities on one dimension can always be obtained across entities on a second dimension for any subset of entities on a third dimension. So, for example, it would be possible (although not necessarily interesting) to isolate classes of stimulus objects from the data and then to obtain correlations between traits and across perceivers for different classes of stimulus objects.

Methodological Problems

Since the range of techniques used in implicit personality theory research is relatively broad, it would be instructive to raise some questions about them. There have, however, been relatively few studies which attempted direct comparisons of methods.

Selection of traits. Although it has been asserted that factorial structure does not vary greatly as a function of stimulus objects and traits (provided a reasonably heterogeneous sample of both is provided), there are reasons to doubt these conclusions. Most studies have provided subjects with both stimulus objects and traits. Correlations among traits differ as a function of both stimulus person (Koltuv, 1962; Veness, 1969) and relevance of trait (Koltuv, 1962). One might suspect that linguisitic factors would predominate in ratings of unfamiliar people using unfamiliar traits.

A second concern with selection procedures also comes in part from Kelly (1955). He suggested that traits were functionally dimensional so that people used both a trait and some opposite. Personal construct theory demands that each trait be paired with its "opposite" to form a construct, but in practice, the opposite or contrast pole is rarely used in research (Bannister & Mair, 1968). In practice, of course, subjects may implicitly use a dimension instead of a single trait in implicit personality theory studies. For example, when asked to judge how likely it is that a *warm* person is also *unhappy*, he may in fact judge how likely a *cold* person is *unhappy* or a *warm* person *happy*, and simply reverse his ratings. Subjects may have difficulty in making direct ratings of negative co-occurrence. Presumably, we are more likely to code those traits which are likely to co-occur with one another rather than those which do not co-occur, but this is an empirical question which deserves further attention.

However, there are apparently real difficulties with the use of opposites. The research of Green and Goldfried (1965) poses the most important problem. They had subjects rate various concepts on several scales. The results indicate that correlations among opposite traits are not highly negative as one would expect. For example, for the concept *sin*, *good* correlated −.39 with *bad* which while negative was far from high. On the other hand, *pleasant* correlated .39 with *unpleasant*, at least when *sin* was being rated. These results then indicate that so-called semantic opposites may typically not behave as opposites in application.

Assignment of traits. Quite obviously, beginning from a Perceiver × Trait × Stimulus Object matrix, correlations among traits can be derived either across perceivers or across stimulus objects. We do not know whether these different ways of calculating correlations would lead to different results. Perhaps even more importantly traits may be distributed across situations rather than people (Mischel, 1968) and may be perceived that way. The results of Endler and Hunt (1966) suggest that situational variance and variance of situational interactions with other variables are important in self-perception, for example.

Comparison of measures of trait relationship. At the level of the matrix of trait relationships, the obvious question is the extent to which different ways of estimating trait relationships correspond.

Todd and Rappoport (1964) have compared several methods. Subjects generated their own traits and then relationships among traits were obtained in three ways: (*a*) Subjects rated 25 stimulus persons on each trait; (*b*) subjects rated idealized concept persons, for example, "intelligent person," on each trait; (*c*) subjects gave co-occurrence ratings. The ratings of stimulus persons were converted into correlations among traits and the co-occurrence ratings were converted to implicating ratings using the Hays (1958) method. The average correlation for the 14 subjects between estimates provided by the two correlational techniques was .83 and between the two correlational and the co-occurrence techniques, .79 and .73. Furthermore, in another part of the study, subjects were asked to pick (*a*) three other traits implied by various other traits and (*b*) what traits would allow them to infer a particular trait. These choices were predicted nicely by the correlational and co-occurrence similarities.

The major advantages of techniques which result in correlations among traits are that the subjects perform a natural judgment task. They assign traits to stimulus objects. Direct techniques on the other hand ask the subject to abstract relationships from his own experience, and of course, for any given perceiver his estimate of how much he sees two traits as related need not agree with the covariance actually present in his judgments.

It might also be pointed out that correlational estimates can provide only symmetrical relationships among traits (e.g., intelligent implies witty as much as witty implies intelligent) while direct measures allow for asymmetricalities which do occur (Schneider[4]; Warr & Knapper, 1968). For example, subjects see witty people as more likely to be intelligent than the reverse (Schneider, see Footnote 4).

Finally, the question of microscopic versus macroscopic analysis has not been sufficiently discussed in this literature. As has been pointed out, researchers aim for a trait relationship matrix which is nearly always reduced to a set of factors or dimensions. Yet, the research of Todd and Rappoport (1964) indicated that while individual trait relationships seemed closely related, factor analysis of the correlational data and a multidimensional scaling of the co-occurrence data gave dimensions which showed poor correspondence. Furthermore, such dimensions did not correspond closely with dimensions subjects provided for themselves. There is suggestive evidence from the research on the importance of the stimulus object to indicate that while the stimulus object is not an important determinant of factorial structure, individual trait relationships may be affected by what objects or persons are being rated.

---

[4] D. J. Schneider. Extra-linguistic aspects of trait implications. In Paul F. Secord, Recent approaches to studying implicit personality theories. Symposium presented at the meeting of the American Psychological Association, Washington, D. C., September 1971.

The evidence on these points is merely suggestive. However, if it is true that factors and dimensions are not congruent with how perceivers think they see the world and factors and dimensions are relatively insensitive to perceiver and stimulus sources of variance, there is reason to consider seriously what levels of analysis are most worthwhile and for what reasons.

OTHER WAYS OF ASSESSING IMPLICIT PERSONALITY THEORY

Stimulus Persons

Several investigators have taken the stimulus person as the unit of investigation. Here the technique is to obtain judgments of similarity among a variety of stimulus persons and then through multidimensional scaling determine stimulus person dimensions. Jackson, Messick, and Solly (1957) had subjects rate the similarity of one another (all subjects were friends of one another), Messick (1961) and Warr and Knapper (1968, Experiment 6) scaled similarity judgments of political figures. Wiggins and Hoffman (1968) had subjects compare profiles representing stimulus persons.

The interpretation of such dimensions can be done in various ways. When the stimulus person has known characteristics (such as with political figures), interpretation was relatively easy. Jackson et al. (1957) obtained personality data on the stimulus persons used in their study so that interpretation in terms of personality traits was possible. Warr and Knapper (1968) showed that the loadings of stimulus persons on various dimensions were related to their ratings on various semantic differential scales. Any dimension(s) on which stimulus persons differ could be used to provide an interpretation of the space of stimulus persons. Wiggins and Hoffman (1968) varied the characteristics of their profiles systematically so that interpretation in terms of any cue dimension or combination of cues was possible through analytic procedures.

It should be pointed out that correlational methods could be adapted to the measurement of stimulus persons. From a Perceiver × Stimulus Person × Trait matrix, correlations could be obtained among stimulus persons across perceivers or traits.

Generally, it seems that scaling of stimulus persons has been an under-used technique. Among other things it allows determination of the dimensions subjects, or groups of subjects, actually use in the judgments of people. Since subjects may be unable or unwilling to verbalize their effective dimensions, person scaling may be one indirect way to assess this.

The Situation

Situations, roles, etc., can also be scaled using either similarity or correlational methods. At present we have little or no empirical evidence about the ways situational forces are perceived by people. While a determination of dimensions along which situations, roles, or other social structures differ is interesting in its own right, it may become an increasingly important problem in person perception. Attribution research has always recognized the role of the situation in perception of behavior (Hastorf, Schneider, & Polefka, 1970). To the extent perceivers see traits and behaviors as distributed across situations or roles, a determination of situation dimensions may be crucial.

STRUCTURE

Because of the interest of personality theorists in cognitive structure during the 1950s and of the general Gestalt impact on person perception it seems natural that much of the early work in implicit personality theory was concerned with structure.

Centrality

Although Asch (1946) did not use the term implicit personality theory, his research in impression formation has profoundly affected theory and methodology in this area. One of his concerns was that of trait centrality: This is discussed at some length by Hastorf et al. (1970).

As is well known, Asch found that *warm–cold* was a central trait in several stimulus contexts. Unfortunately, there was never any good evidence on when a trait would or would not be central or indeed why any other trait should be central at all.

Two separate traditions for considering these problems emerged. The first was based on a correlational model and the second on a model of trait implication. Wishner (1960) was the first to give a correlational basis for a solution. He was able to make a convincing case that the response traits most affected by the shift from *warm* to *cold* were also the traits which correlated most highly with *warm* or *cold*. Thus any trait could become a central stimulus trait if it correlated highly with enough response traits. Warr and Knapper (1968, Experiments 10–13) have replicated Wisher's results closely using narrative material as stimuli.

Most of the research which followed has been animated by a concern with linguistic biases in trait ratings (e.g., D'Andrade, 1965; Kuusinen, 1969a; Mulaik, 1964). It is possible to argue (Jackson, 1962) that centrality is closely related to the loading of a trait on the most powerful factor analytic dimension which is typically evaluative in nature.

A second approach to the questions of method and theory about centrality came from the literature on scaling methods. For Hays (1958), traits were central to the extent that they implied other traits without being implied by them. This judgment can be converted to a measure of distance. Hays (1958) did find that *warm–cold* defined the most prominent dimension in his analysis, but since he used a limited number of traits, his demonstration was merely suggestive.

More recently, a large and careful scaling of 60 traits (Rosenberg et al., 1968) revealed that *warm* and *cold* lie at opposite ends of one important dimension of social evaluation. These authors were also able to show that the traits most strongly affected by *warm–cold* in the initial Asch data loaded heavily on that same dimension. Perhaps just as important, they found that there is also a dimension of intellectual evaluation. These two factors also appeared in a scaling of slang words used to describe Princeton students (Friendly & Glucksberg, 1970).

Rosenberg et al. (1968) argued that traits which are extreme on the two dimensions would be central for traits on their respective dimensions. Zanna and Hamilton (1972) have shown that varying *warm* and *cold* in a list of traits affects only traits which are either strongly socially desirable or undesirable while varying *industrious* and *lazy* affect responses only for intellectually desirable or undesirable traits. These results confirm the Rosenberg et al. hypothesis and confirm the position that centrality may be viewed as a special case of general trait relationships.

In one sense, however, the centrality issue remains an important one. We could ask whether some kinds of traits chosen on an a priori criterion convey more information and imply other traits more strongly than other kinds of traits. For example, do socially desirable traits imply other traits more strongly than negative traits? Hamilton and his colleagues (Hamilton & Huffman, 1971; Hamilton & Zanna [5]) have shown that in judgments of likability and evaluation negative traits are weighted more heavily than positive traits. Clearly more research is needed on the differential power of positive and negative traits.

One might also expect that traits which are more general or abstract might have stronger implications than less general traits. Hinkle (see Footnote 3) had his subjects generate 10 subordinate and 10 superordinate constructs, and he found that the latter were more often involved in causal implication than the subordinate constructs. He also found that the superordinate constructs were more resistant to possible change.

---

[5] D. L. Hamilton & M. P. Zanna. Differential weighting of favorable and unfavorable attributes in impression formation. Unpublished manuscript, Yale University, 1972.

Additive Models

A second issue which came from the Asch paper was an emphasis on additive versus nonadditive processing of information. Basically, he opposed a simple model of traits adding to produce a final impression against a nonpredictive, nonadditive model which claimed that the final impression would be different for each new set of stimuli. It has already been shown that correlational and multidimensional scaling models which are basically additive (although not in the simple way Asch had in mind) can predict final impressions. Furthermore, research from a variety of contexts (Anderson, 1968; Rosenberg et al., 1968; Wiggins, Hoffman, & Taber, 1969) show that linear (additive) models account for almost as much variance as configurational models. Present evidence (e.g., Goldberg, 1968; Wiggins & Hoffman, 1968) indicates that while some people do process configurationally, the configurational models are not dramatically more efficient or accurate. On the other hand, there is some evidence that certain variables may affect trait relationships. Veness (1969) found that occupation affected trait correlations. For example, the correlation between *warm–cold*, and *unhappy–happy* was −.36 for lecturers but .29 for road sweepers (the study was done in England). Over 25% of the correlations differed significantly for different occupations. Further research is likely to be directed to conditions which favor one kind of processing over another (Wiggins & Hoffman, 1968) and more research should be directed to the existence of moderator variables such as occupation.

Cognitive Complexity

A final issue in the structure of implicit personality theory concerns cognitive complexity. At one point it seemed likely that an important aspect of cognitive structure was complexity. Various ways of measuring complexity have been reviewed by Crockett (1965) and Schroder, Driver, and Streufert (1967), among others. Most often complexity has been related to impression formation processes, a literature reviewed by Crockett (1965) and Hastorf et al. (1970).

The main concentration has been on individual differences in complexity. Related directly to implicit personality theory some researchers such as Pedersen (1965) and Walters and Jackson (1966) have found some individual differences in numbers of dimensions used for trait implications, but the differences are subtle and minor. Various researchers such as Steiner (1954) and Warr and Sims (1965) have found evidence that some people rely heavily on an evaluative dimension, thus possibly limiting the number of other dimensions used.

There are other studies which are related to cognitive complexity although they are not generally thought of in this framework. One could ask what features of cognitive set, the situation facing the perceiver, or the kind of stimulus person being judged have on the complexity of implicit personality theory. For our purposes high relationships among traits can be taken as evidence of relatively low complexity, since high relationships probably reflect fewer dimensions or lower independence of dimensions.

The research of Zajonc (1960), Cohen (1961), and others on cognitive tuning are well known. This research showed that information was related in a more complex fashion when subjects expected to receive rather than to communicate the information. More recently O'Neal has shown that cognitive set affects trait relationships. In a first study O'Neal and Mills (1969) argued that when perceivers make a choice among people they desire to be certain and that, therefore, they try to increase the general positive valence of the chosen person. That should result in higher correlations among positive traits for chosen as opposed to unchosen people, a result obtained in the study. In a second study, O'Neal (1971) showed that this halo effect was also increased by caffeine arousal and by decision importance. In both studies only positive traits were used since the interest was in trait covariance due to social desirability. However, it may well be that desires to be certain about an unknown other result in decreased cognitive complexity and a use of fewer dimensions of judgment. When certainty is needed such as when the subject must make a choice or when he is in a communication tuning set, "illusions" of certainty may be achieved through decreased complexity of trait relationships. Thus, trait relationships may be higher for all traits if the above (or other) conditions apply.

There is a possibility that greater complexity (as evidenced by lower trait relationships) may exist for certain types of stimulus persons. Unpublished studies by Nidorf and by Supnick cited by Crockett (1965) showed that information seeking and descriptions are less complex for people who were different from the subject in several strata and who were disliked. Koltuv (1962) found that correlations were higher among traits for unknown than for known others, even when social desirability variance was partialed out. These results are hardly surprising; they are consistent with the position that complexity of implicit personality theory is related to amount of information.

Generally, considerably more attention ought to be addressed to the conditions which increase or decrease the complexity of implicit personality theory. Complexity may be a feature of situations and stimuli as well as perceivers.

REALISM ISSUE

One can always raise the question of where these implicit personality theories came from. If the theory reflects the actual distribution of traits among stimulus persons, then there is an adequate explanation for the theory. Clearly, then, general implicit personality theory would be merely a reflection of the general distribution while individual implicit personality theories represent the effects of idiosyncratic experience.

It is reasonably easy to show that some perceived relationships among traits differ significantly from what is the best guess of an empirical relationship. Thorndike's (1920) halo effect was based on just this kind of observation. This does not, of course, indicate that most perceivers operate with habitual implicit personality theories widely at variance with fact. Not surprisingly, there has been considerable research addressed to this question.

The first explicit attempt to show that implicit personality theory exists independent of subject experiences was by Levy and Dugan (1960). They used 15 traits and their primary measure was how often each was used to describe photographs. There was a separate group of pictures for each trait and no picture was judged on more than one trait. Each perceiver then had a score for each trait reflecting the number of times it was used in rating. Thus they argue correlations among traits could not reflect actual experienced co-occurrence of traits in the sample of photos. Some of the correlations were substantial and a factor analysis produced four major factors, three of which were readily interpretable.

Mulaik (1964) used a more traditional approach. Using the same 20 traits he asked different groups of subjects to rate (*a*) real people or (*b*) stereotyped roles. He also had the 20 traits rated for meaning using a semantic differential technique. Three factors accounting for 60% of the variance in the semantic ratings were quite similar to factors from the first two studies (people judgment). It could therefore be argued that the ratings of people and roles were heavily influenced by linguistic factors.

The question naturally arose as to whether personality ratings could in part be produced by implicit personality theory of the raters and not by actual co-occurrence of traits in the ratees. D'Andrade (1965) showed that ratings of similarity of trait names produced a factor structure close to one arrived at from actual ratings of people. Passini and Norman (1966) had perceivers nominate stimulus persons they hardly knew on a variety of dimensions used earlier by Norman (1963). Factor analyses of such ratings produced a structure which was strikingly similar to one produced by ratings of stimulus persons well known to the raters. Later research by Norman and Goldberg (1966) and Hakel (1969) showed that similar results could be produced in the complete absence of ratees.

Lay and Jackson (1969) also addressed this realism question. They took items which were highly correlated with total scores on 21 major dimensions of the Personality Research Form (Jackson, 1967), a form used to measure Murray needs. Subjects were asked to assess the probabilities of a subject answering one item true, given a true or false answer for another item. Thus distances were generated among all 21 items and a multidimensional scaling analysis done. The first three factors corresponded highly with the first three factors obtained on analysis of subjects' actual test scores. The argument was that since there are correspondences between a measure of real personality and perceived personality, implicit personality theory must in part reflect the real world. Unfortunately, even this conclusion was not completely justified. When subjects filled out the questionnaires, they acted as perceivers of their own behavior; hence the measures of "real" personality may also be biased by implicit personality theory.

Another line of research bears on the realism issue. Chapman (1967) has shown that subjects assume high co-occurrences among stimuli which are linguistically related, a phenomenon he called illusory correlation. Chapman presented pairs of words such that there were no systematic co-occurrences among them, but subjects recalled many more co-occurrences of associatively related words than actually occurred. In later studies, Chapman and Chapman (1967, 1969) found the same phenomenon for the relationship between symptom statements and certain classic signs on the Draw-a-Person Test and for Rorschach signs and judgments of homosexuality. Symptoms which were associatively related to these projective tests were inferred from those signs by naive subjects to the same extent as by trained clinicians. Thus it appeared that inferential relationships among different kinds of information might have been systematically biased such that pieces of information which "should" go together were perceived as in fact co-occurring.

Finally a provocative recent paper by D'Andrade[6] showed that in several studies the matrix of correlations among classes of behaviors rated immediately after social interaction did not correspond highly with the matrix of correlations when the classes of behaviors were rated from memory some time later. The "memory matrix" was, however, closely related to a matrix of linguistic similarity while the "immediate rating matrix" was not. This suggested that trait ratings made from memory were likely to be distorted by feelings of what ought to go together.

The data on the realism issue are not outstanding but what we have suggests that perceived trait relationships deviate from our best estimates of the "true" relationships. Future work is likely to be addressed to the question of what factors bias implicit personality theory, in what direction and in what magnitudes. Such research is likely to discover that there is a variety of situational and cognitive set factors which bias implicit personality theory in addition to cultural and linguistic factors.

LANGUAGE AND IMPLICIT PERSONALITY THEORY

While in principle there may be many "biasing" factors which affect implicit personality theory, in fact, there has been concentration on only one, linguistic determinants. The claim is made and indeed rarely questioned that various kinds of linguistic structures produce the implicit personality theory. It has been shown that semantic differential ratings of trait names produce a linguistic structure highly similar to the structure for other concepts (Kuusinen, 1969b). Other researchers have also obtained factors which are similar to semantic differential factors (Pedersen, 1965), and Rosenberg and Olshan (1970) were able to predict the multidimensional configuration of traits from factors which are quite similar to the Osgood factors.

Most factor analyses of trait names have not produced all of the typical Osgood factors, but there has been a tendency to report a large first factor which has been interpreted as primarily evaluative (Lay & Jackson, 1969; Levy & Dugan, 1960; Pedersen, 1965; Walters & Jackson, 1966). Of course, the early research on halo effect paid particular attention to such factors.

Most recently there has been a substantial controversy about the nature of evaluative bias in person perception. The problem is that in trait ratings there are generally denotative characteristics of the factors in addition to connotative characteristics. For example, Walters and Jackson (1966) noted that their first factor is characterized by high ratings for items concerned with interpersonal warmth and sociability, but it is also the case that all the socially desirable terms load positively and the socially undesirable terms load negatively on this factor. Of course, *warmth* and *sociability* are themselves evaluatively loaded terms so this apparent confounding between evaluation and content is not surprising.

---

[6] R. G. D'Andrade. Cognitive structure and judgment. Paper prepared for the Committee on Basic Research in Education workshop on Cognitive Organization and Psychological Processes, Huntington Beach, California, 1970.

The first researcher to devote considerable attention to separating out these aspects of meaning was Peabody (1967). He organized traits into groups of four. Each group was arranged into a 2×2 block with descriptive and evaluative similarity being the relevant dimensions. Thus each trait (e.g., *cautious*) was associated with a trait which means much the same but with opposite evaluation (e.g., *timid*) and with two traits of opposite descriptive meaning, one positive (e.g., *bold*) and one negative (e.g., *rash*). Subjects were then asked to make inferences from various traits to scales consisting of a positive trait and its negative descriptive opposite (e.g., *cautious–rash* or *timid–bold*). Peabody showed that the overwhelming percentage of inferences were in the direction of descriptive rather than evaluative similarity. That is, subjects inferred that a cautious person would be more likely to be timid than bold. Furthermore, factor analyses of correlations among the individual trait across scales did not produce an evaluative factor although factor analysis of the scales (correlated across traits) revealed factors in which descriptive and evaluative aspects were confounded. Peabody argued that evaluative aspects of meaning were generally secondary to descriptive aspects and that frequently negative traits were more extreme versions of positive ones. For example, the negative *rash* described a more extreme version of boldness than *bold*; similarly, *timid* was a more extreme version of *cautious*.

Recently, Peabody's research has been criticized on two counts. Rosenberg and Olshan (1970) argued that the factor analyses based on correlations of traits across scales inadvertently discarded evaluative variance. Such a correlation reflected only similarities in shapes of profiles and discarded differences in means for traits. These authors found that a reanalysis of Peabody's data based on profile distances provided dimensions which were in part evaluative. In a reanalysis of his earlier data Peabody (1970) concurred that the original study underestimated evaluative effects at the dimensional level.

A second set of criticisms was provided by Felipe (1970). He argued that Peabody's traits were highly related descriptively and hence the evaluative dimension may have been submerged. He showed that indeed the evaluative similarities seemed to appear more critical and descriptive similarity less important than Peabody had thought. Peabody (1970) in reply had modified his theory to suggest that in relationships between traits evaluative effects are likely to be constant while descriptive effects vary. If they are stronger than evaluative effects, they will seem to predominate.

A somewhat different approach to the whole question of evaluative bias came from the work of Kuusinen (1969a). He used a simple but elegant technique to partial out evaluative (and also potency and activity) aspects of trait ratings. The resultant factor analysis produced factors which differed in significant ways from the unpartialed structure. The author claimed that the unpartialed matrix had greater psychological significance.

Arguments about linguistic determination of implicit personality theory could continue forever. As argued earlier in this paper, implicit personality theory related to known others and using relevant traits might be less influenced by linguistic factors than studies using unfamiliar stimulus persons and traits. This problem has not been studied explicitly.

INDIVIDUAL DIFFERENCES

The claim has been made that the search for personality correlates of implicit personality theory is bound to fail since there are few individual differences among perceivers in implicit personality theory (Kuusinen, 1969b). This claim rests on the interpretation of factor analysis data, where Kuusinen attempted to show that the congruence among factors derived from individual subjects was high.

This conclusion was disputed by Hamilton (1970) as being based on a too lenient criterion of what constitutes factor congruence. Also Hamilton argued that the Kuusinen data gave evidence of both individual differences and commonality. Using a more stringent (and more widely accepted criterion), Hamilton showed that there are extensive individual differences, both in the Kuusinen data and in new data collected by Hamilton.

The research on individual differences can be divided into two groups. Some researchers have isolated individual differences in implicit personality theory and then have tried to assess how these groups of perceivers differ. A second group has begun with a personality variable and has attempted to see whether people who differed along the personality dimension also varied in their perceptions of others.

Those studies which have used the first general procedure have not been noticeably successful. The most common technique has been to use an individual differences model of multidimensional scaling. This model allows one to separate out of the trait implication matrix groups of subjects who differ in their distances among traits. While theoretically powerful, in practice the result is to find a major point of view with one or more groups of deviant subjects who differ in relatively small ways from the first group. Furthermore, the points of view do not seem to be strongly related to personality variables.

Walters and Jackson (1966) found two points of view with one being generally characterized as making less extreme judgments than the other. There were almost no personality correlates. Pedersen (1965) located three points of view and one point of view seemed to be related to authoritarianism. Messick and Kogan (1966) found several points of view, some of which seemed related to broad versus narrow category use, but again, the personality data were weak. For the moment it appears that classic personality variables are not strongly related to the different points of view isolated in multidimensional scaling.

Some studies have relied on correlational analyses to isolate individual differences. Posavac (1971) presented subjects with pairs of traits and had them indicate their preferences. From the Subject × Trait-Pair matrix correlations among subjects were calculated and seven subject factors isolated. Interpretations for each subject factor were made in terms of personality data. While the correlations were quite low, interpretations were possible for some dimensions. Hamilton and Yellig[7] did factor analyses for separate subjects and they found not only low congruence among subjects but high congruence for each subject over time. Thus it would seem that the fact that personality variables are not related to individual differences is not a function of reliable dimension determination.

There is a second way to approach these problems. One might select a priori groups and see how their perceptions differ. Warr and Knapper (1968, Experiment 1) and Messick (1956, 1961) have used this approach with mixed success.

At a more microscopic level there are several reports of individual differences in trait inferences. For example, Steiner (1954) and Warr and Sims (1965) found that high authoritarians make more extreme inferences from traits which are extremely positive or negative. Jones (1954) found that high authoritarians are highly responsive to power cues in stimulus persons. Warr and Knapper (1968, Experiment 14) and Benedetti and Hill (1960) showed that personality differences relate to extremeness of inference for some cue traits. Shapiro and Tagiuri (1959), Warr and Knapper (1968, Experiment 14), and Lay (1970) have all found that women are more inclined to make extreme inferences from certain stimuli than men. Matkom (1963) found that maladjusted subjects are more certain with positive information than adjusted subjects. Finally, Secord and Berscheid (1963) found minimal differences between prejudiced and unprejudiced subjects in their attributions from stereotyped traits to black and white stimulus persons.

It is relatively easy to show that individuals differ in their implicit personality theory, but there has been limited success relating such differences to traditional personality variables. This may, of course, be more damaging to traditional personality variables than implicit personality theory. Future research might concentrate more on analyses of single perceivers, an approach used ingeniously by Rosenberg and Jones (1972) to study the novelist Theodore Dreiser. Greater attention might also be given to differences among perceivers in what categories they use or prefer to use to judge others, a literature reviewed in the next section.

---

[7] Reported in D. L. Hamilton. Implicit personality theories: Dimensions of interpersonal cognition. In Paul F. Secord, Recent approaches to studying implicit personality theory. Symposium presented at the meeting of the American Psychological Association, Washington, D. C., September 1971.

CONTENT QUESTIONS

Nontrait Aspects of Implicit Personality Theory

Most of the research reviewed thus far has concerned the relationship of traits or, occasionally, attitudes. There is no reason why any attribute of persons can not be used in an implicit personality theory paradigm. Indeed, as will be argued presently, this would more than likely broaden the conception of implicit personality theory considerably.

There have been several lines of research which have not generally been thought of in implicit personality theory terms but which obviously fit the paradigm.

Facial features. One main tradition concerns the inferential relationships between facial features and traits and is best represented by the work of Secord (1958). There was considerable agreement among perceivers on which traits they were willing to attribute to particular photographs, and variables such as darkness of skin, the use of cosmetics, and curvature of the mouth led directly to personality inferences. There was also some evidence that subjects inferred certain physical traits from personality descriptions.

Emotions. A second line of research is usually subsumed under the topic of person perception accuracy, and it relates to the whole question of inferring momentary states such as emotion from physical, primarily facial cues. Although the emphasis has tended to be on the accuracy of inference and cross-cultural constancies in emotional expression and perception, there is no reason why such an approach could not be used to study implicit personality theory. Another variety of such research is represented by work on the scaling of facial expressions (Frijda, 1969). Since dimensions of such expressions are generally interpreted in terms of personality and emotional attributes, the work has some relevance to implicit personality theory.

Stereotypes. Another kind of research which has an obvious relationship to implicit personality theory is the work in stereotyping. As Hastorf et al. (1970) indicated, stereotypes were one kind of implicit personality theory and, indeed, race could be viewed as a classic central trait for socalled stereotyped traits. A recent review of this literature (Brigham, 1971) indicated that race is a powerful cue in trait attributions. One important question is whether the structure of trait relationships is different for say black and white stimulus persons. Secord and Berscheid (1963) found no evidence of this.

Other attributes. It should be indicated that there have been relatively few studies on the inferences made from socioeconomic variables, physical attractiveness, style of dress, etc. Gibbins (1969) did find that style of dress led to inferences about people. Shyness, occupations, number of boyfriends, and smoking were among the characteristics differentially attributed to wearers of particular styles of clothing. Expensive suits, attractive figures, beards, and love beads seemed to lead to powerful inferences in everyday life.

Finally, there have been few studies on the relationships between behaviors and traits. Attribution research (Hastorf et al., 1970) has been concerned primarily with conditions which allow one to separate situational and personal causality, but there may simply be behaviors which are thought to be "revealing" in and of themselves. Nail biting implies nervousness, for example, at least for many perceivers. And, of course, classic psychoanalytic personality theories have suggested classes of behavior, which ought to imply certain broad personality syndromes. Again, there has been no research attention to relatively stable inferential relationships between behaviors and traits.

Units of Person Cognition

The fact that most studies of implicit personality theory have used traits may be largely an historical accident resulting from the importance of traits in personality theory and personality assessment research, and from the fact that much of the early research in the field was stimulated by Asch's impression formation research where traits had been used merely as a convenience. The fact of the matter is, we have little or no hard evidence on what the natural units of person cognition are.

A more general point is that we have, at present, no basis for determining the circumstances under which perceivers prefer one kind of trait to another or for that matter whether and when perceivers feel comfortable with traits as opposed to more behavioral or typological units. Given that a person can be described behaviorally ("He can't look another in the eye"), in trait language ("He's shy"), or in a language of typology ("He's introverted"), when do we choose to use a particular level? While attribution research addresses this question, it could be argued that so long as such research provides perceivers with categories of explanation, it can never answer the question of which categories are "naturally" used by perceivers. Attribution research as well as implicit personality theory would be improved by an analysis of what units people use under different circumstances. It is worth noting that Bruner and Tagiuri made the same plea in 1954.

Category Relevance

An issue which seems to pop up in 10-year cycles is that of trait relevance. It has been argued (Hastorf, Richardson, & Dornbush, 1958; Kelly, 1955; MacLeod, 1947) that results may be biased in unknown ways when subjects are forced to use experimenter-provided categories.

One of the major problems faced in such research is how to code subjects' responses. Kelly (1955), of course, had subjects make responses in terms of traits and their opposites. Rosenberg (Rosenberg & Sedlak, 1972) also had subjects respond in discrete units, and some reduction into trait categories was used. Dornbush, Hastorf, Richardson, Muzzey, and Vreeland (1965), Beach and Wertheimer (1961), and Yarrow and Campbell (1963) have all developed coding categories.

There are several questions involved with relevance. The first question is whether the relevance of categories or traits is more a stable set which the perceiver brings to any person description task or whether it varies with the stimulus person and the situation. Dornbush et al. (1965) reported that a given perceiver more likely uses the same categories in assessing two stimulus persons than two perceivers describing the same stimulus person. Yarrow and Campbell (1963) more informally reported similar results. On the other hand, Beach and Wertheimer (1961) reported low reliability of category usage in a test–retest measure; in the retest, however, subjects could use different stimulus persons and many did.

A closely related question concerns the reliability of categories, constructs, and traits. Yarrow and Campbell (1963) reported low reliabilities of categories over time. Fjeld and Landfield (1961) have shown high repeat use of constructs for the same stimulus persons over a 2-week period. Mitsos (1958) found that if the same role figures were used, the same constructs were used over a 3-month period, but the reliability was lessened if different stimulus persons were used in the second administration.

A third question—whether categories vary with perceiver and stimulus person characteristics. Yarrow and Campbell (1963), Beach and Wertheimer (1961), and Richardson, Hastorf, and Dornbush (1964) all report differential category usage by various groups of people for different kinds of stimulus persons and in some cases for interactions between these variables. At present, there seems to be no general way to summarize these results.

A fourth question is whether relevant traits are used differently than nonrelevant traits. Koltuv (1962) showed that subjects assumed higher relationships among more relevant than among less relevant traits. A variety of explanations is possible for this effect, but given the finding of Warr and Knapper (1968, Experiment 3) that test–retest reliability of scales was positively related to how applicable the scales were to a given stimulus person, the results of Koltuv may simply reflect the greater reliability of relevant scales. Cromwell and Caldwell (1962), Mitsos (1961), Landfield (1968), and Tajfel and Wilkes (1963) have also shown that subjects made more extreme judgments on constructs they provided for themselves. The effect was not replicated by Warr and Coffman (1970). If, however, the effect was generally true, the Koltuv results could also be explained in terms of greater variance for the relevant scales.

The work of Wiggins et al. (1969) and Hamilton and Gifford (1970) can also be thought of in a relevance framework. These researchers presented subjects with profiles which differed along several dimensions. Subjects rated each profile on some criterion or rated the similarity of each pair of profiles (Wiggins & Hoffman, 1968), and it was possible to determine which dimensions relate most closely to the profile ratings. Thus it was possible to determine empirically which of several dimensions subjects or any group of subjects actually used. It was also possible with this technique to isolate groups of perceivers who weight cues in the same manner and determine how these groups of perceivers differ (Wiggins et al., 1969). This technique is potentially a powerful one for isolating cues which are relevant not in the sense of being used in descriptions but in the sense of being informative for different perceivers.

CONCLUSION

Several conclusions emerge from this review. First, researchers have not availed themselves of all the possible techniques for assessing implicit personality theory. Second, there has been an emphasis on the dimensional aspects of trait similarities to the relative exclusion of questions concerning the content and dynamic qualities of implicit personality theory. The sophistication of methods is greater than the sophistication of the substantive questions.

REFERENCES

Anderson, N. H. A simple model for information integration. In R. Abelson, E. Aronson, W. McGuire, T. Newcomb, M. J. Rosenberg, and P. Tannenbaum (Eds.), *Theories of cognitive consistency*. Rand-McNally, 1968.

Asch, J. E. Forming impressions of personalities. *Journal of Abnormal and Social Psychology*, 1946, **41**, 258–290.

Bannister, D., & Mair, J. M. M. *The evaluation of personal constructs*. London: Academic Press, 1968.

Beach, L., & Wertheimer, M. A free response approach to the study of person cognition. *Journal of Abnormal and Social Psychology*, 1961, **62**, 367–374.

Benedetti, D. T., & Hill, J. G. A determiner of the centrality of a trait in impression formation. *Journal of Abnormal and Social Psychology*, 1960, **60**, 278–280.

Brigham, J. C. Ethnic stereotypes. *Psychological Bulletin*, 1971, **76**, 15–38.

Bruner, J. S., Shapiro, D., & Tagiuri, R. The meaning of traits in isolation and in combination. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford University Press, 1958.

Bruner, J. S., & Tagiuri, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 2. Cambridge, Mass.: Addison Wesley, 1954.

Chapman, L. J. Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 151–155.

Chapman, L., & Chapman, J. P. The genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, 1967, **72**, 193–204.

Chapman, J. J., & Chapman, J. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 1969, **74**, 271–280.

Cohen, A. R. Cognitive tuning as a factor affecting impression formation. *Journal of Personality*, 1961, **29**, 235–245.

Crockett, W. H. Cognitive complexity and impression formation. In B. A. Maher (Ed.), *Progress in experimental personality research*. Vol. 2. New York: Academic Press, 1965.

Cromwell, R. L., & Caldwell, D. F. A comparison of ratings based on personal constructs of self and others. *Journal of Clinical Psychology*, 1962, **18**, 43–46.

Cronbach, L. J. Processes affecting scores on understanding of others and assuming "similarity." *Psychological Bulletin*, 1955, **52**, 177–193.

Cronbach, L. J., & Gleser, G. C. Assessing similarity between profiles. *Psychological Bulletin*, 1953, **50**, 456–473.

D'Andrade, R. G. Trait psychology and componental analysis. *American Anthropologist*, 1965, **67**, 215–228.

Dornbush, S., Hastorf, A. H., Richardson, S. A., Muzzy, R. E., & Vreeland, R. S. The perceiver and the perceived: Their relative influence on the categories of interpersonal perception. *Journal of Personality and Social Psychology*, 1965, **1**, 434–440.

Endler, N. S., & Hunt, J. McV. Sources of behavioral variance as measured by the S–R Inventory of Anxiousness. *Psychological Bulletin*, 1966, **65**, 336–346.

Felipe, A. I. Evaluative versus descriptive consistency in trait inferences. *Journal of Personality and Social Psychology*, 1970, **16**, 627–638.

Friendly, M. L., & Glucksberg, S. On the description of subcultural lexicons: A multidimensional approach. *Journal of Personality and Social Psychology,* 1970, **14**, 55–65.

Frijda, N. H. Recognition of emotion. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 4. Academic Press, 1969.

Fjeld, S. P., & Landfield, A. W. Personal construct consistency. *Psychological Reports,* 1961, **8**, 127–129.

Gibbins, K. Communication aspects of women's clothes and their relation to fashionability. *British Journal of Social and Clinical Psychology,* 1969, **8**, 301–312.

Goldberg, L. R. Simple models or simple processes? Some research on clinical judgments. *American Psychologist,* 1968, **23**, 483–496.

Green, R. F., & Goldfried, M. R. On the bipolarity of semantic space. *Psychological Monographs,* 1965, **79**(6, Whole No. 599).

Gross, C. Intrajudge consistency in ratings of heterogeneous persons. *Journal of Abnormal and Social Psychology,* 1961, **62**, 605–610.

Hakel, M. Significance of implicit personality theories in personality research and theory. *Proceedings of the 77th Convention of the American Psychological Association,* 1969, **4**, 403–404.

Hamilton, D. L. The structure of personality judgments: Comments on Kuusinen's paper and further evidence. *Scandanavian Journal of Psychology,* 1970, **11**, 261–265.

Hamilton, D. C., & Gifford, R. K. Influence of implicit personality theories on cue utilization in interpersonal judgment. *Proceedings of the 78th Annual Convention of the American Psychological Association,* 1970, **5**, 415–416.

Hamilton, D. L., & Huffman, L. J. Generality of impression-formation processes for evaluative and nonevaluative judgments. *Journal of Personality and Social Psychology,* 1971, **20**, 200–207.

Hastorf, A. H., Richardson, S. A., & Dornbush, S. M. The problem of relevance in the study of person perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford: Stanford University Press, 1958.

Hastorf, A. H., Schneider, D. J., & Polefka, J. *Person perception.* Reading, Mass.: Addison-Wesley, 1970.

Hays, W. L. An approach to the study of trait implication and trait similarity. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford: Stanford University Press, 1958.

Jackson, D. N. The measurement of perceived personality trait relationships. In N. Washburn (Ed.), *Values, decisions and groups.* Vol. 2. New York: Pergamon, 1962.

Jackson, D. N. *Personality research form manual.* Goshen, N. Y.: Research Psychologists Press, 1967.

Jackson, D. N., Messick, S., & Solley, C. M. A multidimensional scaling approach to the perception of personality. *Journal of Psychology,* 1957, **44**, 311–318.

Johnson, S. C. Hierarchical clustering schemes. *Psychometrika,* 1967, **32**, 241–254.

Jones, E. E. Authoritarianism as a determinant of first-impression formation. *Journal of Personality,* 1954, **23**, 107–127.

Kelly, G. A. *The psychology of personal constructs.* New York: Norton, 1955.

Koltuv, B. Some characteristics of intrajudge trait intercorrelations. *Psychological Monographs,* 1962, **76**(33, Whole No. 552).

Kuusinen, J. Affective and denotative structures of personality ratings. *Journal of Personality and Social Psychology,* 1969, **12**, 181–188. (a)

Kuusinen, J. Factorial invariance of personality ratings. *Scandanavian Journal of Psychology,* 1969, **10**, 33–44. (b)

Landfield, A. W. The extremity rating revisited within the context of personal construct theory: *British Journal of Social and Clinical Psychology,* 1968, **7**, 135–139.

Lay, C. H. Trait-inferential relationships and judgments about the personalities of others. *Canadian Journal of Behavioral Sciences,* 1970, **2**, 1–17.

Lay, C. H., & Jackson, D. N. Analysis of the generality of trait-inferential relationships. *Journal of Personality and Social Psychology,* 1969, **12**, 12–21.

Levy, L. H., & Dugan, R. D. A constant error approach to the study of dimensions of social perception. *Journal of Abnormal and Social Psychology,* 1960, **61**, 21–24.

MacLeod, R. The phenomonological approach to social psychology. *Psychological Review,* 1947, **54**, 193–210.

Matkom, A. J. Impression formation as a function of adjustment. *Psychological Monographs,* 1963, **77**(5, Whole No. 568).

Messick, S. J. The perception of social attitudes. *Journal of Abnormal and Social Psychology,* 1956, **52**, 57–66.

Messick, S. J. The perceived structure of political relationships. *Sociometry,* 1961, **24**, 270–278.

Messick, S. J., & Kogan, N. Personality consistencies in judgment: Dimensions of role constructs. *Multivariate Behavioral Research,* 1966, **1**, 165–175.

Mischel, W. *Personality and assessment.* New York: Wiley, 1968.

Mitsos, S. B. Representative elements in role construct technique. *Journal of Consulting Psychology,* 1958, **22**, 311–313.

Mitsos, S. B. Personal constructs and the semantic differential. *Journal of Abnormal and Social Psychology,* 1961, **62**, 433–434.

Mulaik, S. A. Are personality factors raters' conceptual factors? *Journal of Consulting and Clinical Psychology,* 1964, **28**, 506–511.

Newcomb, T. An experiment designed to test the validity of a rating technique. *Journal of Educational Psychology,* 1931, **22**, 279–289.

Norman, W. T. Toward an adequate taxonomy of personality attribute: Replicated factor structure in peer nomination personality ratings. *Journal of*

Abnormal and Social Psychology, 1963, **66**, 574–583.

Norman, W. T., & Goldberg, L. R. Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology*, 1966, **4**, 681–691.

O'Neal, E. Influence of future choice importance and arousal upon the halo effect. *Journal of Personality and Social Psychology*, 1971, **19**, 334–340.

O'Neal, E., & Mills, J. The influence of anticipated choice on the halo effect. *Journal of Experimental Social Psychology*, 1969, **5**, 347–351.

Osgood, C. E. Studies on the generality of affective meaning systems. *American Psychologist*, 1962, **17**, 10–28.

Passini, F. T., & Norman, W. T. A universal conception of personality structure? *Journal of Personality and Social Psychology*, 1966, **4**, 44–49.

Peabody, D. Trait inferences: Evaluative and descriptive aspects. *Journal of Personality and Social Psychology Monograph*, 1967, **7**(4, Pt. 2).

Peabody, D. Evaluative and descriptive aspects in personality perception: A reappraisal. *Journal of Personality and Social Psychology*, 1970, **16**, 639–646.

Pedersen, D. M. The measurement of individual differences in perceived personality-trait relationships and their relation to certain determinants. *Journal of Social Psychology*, 1965, **65**, 233–258.

Posavac, E. J. Dimensions of trait preferences and personality type. *Journal of Personality and Social Psychology*, 1971, **19**, 274–281.

Richardson, S. A., Hastorf, A. H., & Dornbush, J. M. Effects of physical disability on a child's description of himself. *Child Development*, 1964, **35**, 893–907.

Rosenberg, S., & Jones, R. A method for investigating and representing a person's implicit theory of personality: Theodore Dreiser's view of people. *Journal of Personality and Social Psychology*, 1972, **22**, 372–386.

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, 1968, **9**, 283–294.

Rosenberg, S., & Olshan, K. Evaluative and descriptive aspects in personality perception. *Journal of Personality and Social Psychology*, 1970, **16**, 619–626.

Rosenberg, S., & Sedlak, A. Structural representations of implicit personality theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 6. New York: Academic Press, 1972.

Schroder, H. M., Driver, M. J., & Streufert, S. *Human information processing*. New York: Holt, Rinehart & Winston, 1967.

Secord, P. Facial features and inference processes in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford University Press, 1958.

Secord, P., & Berscheid, E. Stereotyping and the generality of implicit personality theory. *Journal of Personality*, 1963, **31**, 65–78.

Shapiro, D., & Tagiuri, R. Sex differences in inferring personality traits. *The Journal of Psychology*, 1959, **47**, 127–136.

Steiner, I. D. Ethnocentrism and tolerance of trait inconsistency. *Journal of Abnormal and Social Psychology*, 1954, **49**, 349–354.

Tajfel, H., & Wilkes, A. L. Salience of attributes and commitment to extreme judgments in the perception of people. *British Journal of Social and Clinical Psychology*, 1963, **2**, 40–49.

Thorndike, E. L. A constant error in psychological ratings. *Journal of Applied Psychology*, 1920, **4**, 25–29.

Todd, F. J., & Rappoport, L. A cognitive structure approach to person perception: A comparison of two models. *Journal of Abnormal and Social Psychology*, 1964, **68**, 469–478.

Veness, T. Trait assessment intercorrelation and occupational stereotypy. *British Journal of Social and Clinical Psychology*, 1969, **8**, 227–234.

Walters, H. A., & Jackson, D. N. Group and individual regularities in trait inference: A multidimensional scaling analysis. *Multivariate Behavioral Research*, 1966, **1**, 145–163.

Warr, P. B., & Coffman, T. L. Personality, involvement and extremity of judgment. *British Journal of Social and Clinical Psychology*, 1970, **9**, 108–121.

Warr, P. B., & Knapper, C. *The perception of people and events*. New York: Wiley, 1968.

Warr, P. B., & Sims, A. A. A study of cojudgment processes. *Journal of Personality*, 1965, **33**, 598–604.

Wiggins, N., & Hoffman, P. J. Three models of clinical judgment. *Journal of Abnormal Psychology*, 1968, **73**, 70–77.

Wiggins, N., Hoffman, P. J., & Taber, T. Types of judges and cue utilization in judgments of intelligence. *Journal of Personality and Social Psychology*, 1969, **12**, 52–59.

Wishner, J. Reanalysis of "Impressions of personality." *Psychological Review*, 1960, **67**, 96–112.

Yarrow, M. R., & Campbell, J. D. Person perception in children. *Merrill-Palmer Quarterly*, 1963, **9**, 57–72.

Zajonc, R. B. The process of cognitive tuning in communication. *Journal of Abnormal and Social Psychology*, 1960, **61**, 159–167.

Zanna, M. P., & Hamilton, D. L. Attribute dimensions and patterns of trait inferences. *Psychonomic Science*, 1972, **27**, 353–354.

(Received April 10, 1972)

Psychological Bulletin
1973, Vol. 79, No. 5, 310–317

REACTANCE AND THE UNWILLINGNESS TO HELP OTHERS[1]

LEONARD BERKOWITZ[2]

University of Wisconsin

A demand, explicit or implicit, to help someone, and even a felt obligation to do this, is often resented because the demand or obligation is a bothersome threat to the individual's freedom of action. Evidence is cited in accord with Brehm's reactance theory indicating that increased pressure to aid a person in need at times reduces the individual's willingness to help the person who is dependent upon him. Experiments demonstrate that this "reactance effect" is lessened when the individual is in a good mood, and is increased when he is self-concerned and when the help request seems improper or unwarranted. People might be induced to act helpfully in spite of the threat to their independence by procedures that clearly establish helpfulness as socially desirable conduct and provide practice in behaving helpfully.

In March, 1964, Catherine Genovese was stabbed to death in three separate attacks lasting more than half an hour even though thirty-eight respectable, law-abiding citizens knew of her plight. Not one person telephoned the police during the assault; one witness called after the woman was dead. Commenting on the incident later, Rosenthal (1964) attributed the failure to help Miss Genovese to "a widespread indifference." Only under certain situations, he believed, would "a man step out of his shell toward his brother [p. 81]."

We can offer a very general and quite obvious reason why man doesn't readily help his brother: The effort entails some cost, real or anticipated, and he is often unwilling to incur this cost. But to say this much is clearly not enough. We must determine what kinds of costs are involved in helping other people, and also explain why help is at times given in spite of these costs.

The well-known experiment by Darley and Latané (1968) illustrates the reluctance to help someone in need even when the effort required is not too great. The undergraduates in this study, supposedly assembled in groups of two, three, or six members were slower to call for help when one of their discussion partners had a (staged) epileptic seizure, the larger the number of other bystanders in the group. In other words, there evidently was a stronger tendency to "let George do it" the more Georges there were in the situation who might also help the victim. The investigators speculated that the subjects felt a greater diffusion of responsibility with the increased number of supposed bystanders who could also do something.[3]

Why were the undergraduates unwilling to aid the other person? They apparently recognized an obligation to help him, but even so, many of the subjects did not want to come to his assistance, and rationalized away their obligation if other people were present. The other bystanders provided a welcome alibi.

The subjects could have been afraid of a number of things. As Latané and Darley (1970) noted, the students might have been reluctant to intrude into someone else's personal affairs. The stranger's epilepsy was his private problem and they did not want to embarrass him. They probably also did not want to act inappropriately in this strange situation. Not knowing what to do, they would do nothing if other people were around. Supporting this possibility, subjects in a later experiment by Schwartz and Clausen (1970) reported the emergency much more quickly if they knew exactly how to help the victim than if they did not have this information. Furthermore, as Latané and Darley (1970) also suggested, the subjects could also have been wary of the legal entanglements that might have arisen if they took action. All of these fears undoubtedly existed. Latané and Darley also found that subjects were much quicker to aid the epilepsy victim if the other bystander was a friend rather than a stranger, perhaps because the friend's presence helped alleviate some of these anxieties.

---

[1] The author's research and the preparation of this paper were aided by Grant GS-1890X from the National Science Foundation.

[2] Requests for reprints should be sent to Leonard Berkowitz, Department of Psychology, University of Wisconsin, Madison, Wisconsin.

[3] In a later replication of this experiment Schwartz and Clausen (1970) obtained a significant decrease in speed of helping with an increased number of bystanders only in the case of females rather than males. The findings in this later study also indicated that the addition of bystanders affected only the reporting of the emergency to others and not the subject's likelihood of taking direct action.

This situation may also arouse another type of motive: a desire not to get involved with someone else because the involvement lessens one's freedom of action. Going to the victim's aid might mean having to stay with him for some time. Perhaps it would require a continued commitment to help him. More generally, the victim has some control over his potential helper to the extent that this latter person feels pressure to aid the individual in need,[4] and the subjects might even have resented the obligation they felt to help the victim. Responsibility for another person's welfare is also a pressure threatening one's freedom.

This resentment can be important when someone needs help. Whatever other costs might be anticipated in such situations—the risk of bodily harm, unwanted muscular effort, the fear of intruding, and so on—many people are often unwilling to assist someone because his need for assistance also implies a bothersome restriction of their independence. This reasoning closely parallels Brehm's (1966) theory of psychological reactance. Brehm maintains that "reactance" is generated by threats to the individual's behavioral freedom. The result, Brehm says, is hostility, as well as an increased desire to do whatever it is that the individual believes he may not be able to do. In general, then, reactance is aroused when a person learns that someone is dependent upon him for help, and this reactance could lessen the potential helper's willingness to aid the dependent individual.

EVIDENCE OF REACTANCE

An experiment by Latané and Darley (1970, pp. 81–85) can be interpreted as providing evidence for reactance. Male college students who were waiting to be interviewed for an opinion survey individually overheard a fight between two children in the adjoining room. A bully seemed to be picking on a much younger child. In one condition the subjects were led to think the children were alone, while in another group they had been informed that another adult was in the room with the youngsters. The researchers assumed that the subject would feel some responsibility to stop the fight if he were the only adult around, and would think himself less obliged to intervene if the other adult in the room with the children could stop the fight. The felt pressure to intervene in the former condition apparently was unpleasant and motivated the men in this group to deny their responsibility. When they were later asked to say what had been going on in the next room, fully 88% of the men in the condition having another adult in the room with the children said the youngsters had been having a real fight. On the other hand, if they had been the only adult present, so their responsibility was great, only 25% of them said the fight was real. The felt responsibility was an unwelcome burden, leading most of the subjects to deny it.

Reactance can also be seen in an experiment by Berkowitz, Klanderman, and Harris (1964). The subjects were to construct paper boxes in behalf of a peer so that the greater the number of boxes they produced, the more help they gave him, although they did not stand to gain any explicit benefits for themselves by working hard. In half of the cases, the experimenter would see their output right away, while the remaining subjects were told that the experimenter would not learn how much help they had given their peer (because their work was to be placed inside a sealed carton) until the end of the semester weeks away. The subjects were either men or women and the experimenter was also either male or female.

Analysis of the increase in productivity from a prior practice period revealed a significant interaction among the three variables, sex of subject, sex of experimenter, and how soon the experimenter would learn of their effort. When confronted by an opposite-sex experimenter, the subjects tended to work harder if the experimenter was immediately aware of their effort than if he was not. They wanted to gain the opposite-sex experimenter's approval if it could be gotten quickly. However, if the experimenter was of the same sex, they actually worked less hard when the experimenter could see their effort at that time than when their work would not become apparent for several weeeks This same-sex experimenter's immediate awareness of their work was an additional pressure which evidently lessened their willingness to aid their partner. Instead of complying to the watching experimenter's "demand," they resisted it, presumably because of the reactance they felt.[5]

---

[4] Schopler and Bateson (1965) phrased this nicely when they entitled a study of the tendency to help dependent others "the power of dependence."

In this experiment the experimenter's demand was only implied. Jackson has shown that the experimenter's explicit request can also generate reactance which might lessen the subjects' motivation to help him out. In Jackson's[6] study the experimenter first did half of the subjects a favor following a procedure to be described later in this paper—he voluntarily provided each person in this condition with a soft drink. Then, some subjects were told that the experimenter was highly dependent upon them because he greatly needed "good" results from them, while other subjects were led to think he was much less dependent upon them. The performance was a verbal conditioning task. Jackson found that subjects who were aware of the reinforcement contingency displayed more verbal conditioning the greater the experimenter's dependency upon them—but only if the experimenter had not done them the favor earlier. Presumably because the prior favor had increased the pressure they felt, the subjects given the soft drink earlier actually showed *less* verbal conditioning under high than low experimenter dependence.

As was suggested earlier, reactance might also be generated by the possibility of a longstanding commitment to help someone. The threat of someone's persistent control over the individual could be particularly bothersome. Jones (1970) found that college men were least willing to aid a dependent peer when his request implied that he would need their services for some time in the future.

The reactance conception implies, of course, that resentment should increase as greater pressure is placed on the individual. More specifically in regard to helpfulness, the stronger the demand, explicit or implicit, that he aid someone, the more intense should be the reactance generated. But greater pressure should also heighten the chances of compliance (the individual will feel more obligated to do the "right" thing) up to some maximum. Beyond this point the individual might well rebel so that extremely strong pressure actually produces decreased helpfulness, especially if he can find some justification for not giving assistance.[7]

Schwartz (1970) asked people who had just given blood at a local Red Cross center to join a pool of possible bone marrow donors waiting to aid a young woman. The implicit pressure was manipulated in two ways: One, by mentioning the possible consequences for this woman—from a statement that she might be helped (low pressure), through a middle level, to a third, most intense level created by the statement that she was unlikely to survive without a matching bone marrow donation and that her death would be a tragedy for her children. Second, the demand was also varied by saying the odds were either 1 in 1,000 or 1 in 25 that the subject's own bone marrow would match the woman's and that he would actually be asked to be a donor. When the odds were quite small that the subject would be a suitable donor, the greater the pressure produced by emphasizing the consequences for the young woman, the more willingly the subjects agreed to join the pool of possible donors. However, if there was a fairly good chance that the person would actually be called upon, there was a curvilinear relationship between the consequences manipulation and the subject's willingness to volunteer; the subjects were actually most reluctant to be a possible donor when they had been exposed to the greatest pressure—telling them that the woman probably would die without a successful transplant.

---

[5] Moscovici and Neve (1971) found, in accord with these results, the subjects were more likely to conform to another's expressed judgments when that person was out of the room as compared to when she was present.

[6] J. S. Jackson. The impact of a favor and dependency of the reinforcing agent on social reinforcer effectiveness. Unpublished doctoral dissertation, Wayne State University, 1972.

[7] The consultant who had read this paper for the journal advanced the interesting possibility that "the positive function relating pressure to motivation to help is negatively accelerating" while "the positive function relating pressure to reactance" may be "positively accelerating."

CONDITIONS AFFECTING TOLERANCE OF REACTANCE

The following discussion assumes that reactance is typically aroused when people learn that others need their aid, but that various situational conditions affect their tolerance of this reactance. Although there are some data in accord with this view (Berkowitz, 1969), it is also quite possible that these conditions influence the experienced magnitude of the reactance rather than the person's capacity to withstand it.

Whichever alternative is correct, people are usually much more willing to aid the individual requiring assistance because of factors beyond his control than the person whose dependency is his own fault (Schopler & Matthews, 1965). They evidently are often unhappy about the obligation imposed on them —or the obligation they feel—to help the individual in need, but may not actively resist the felt demand unless they regard the demand as unwarranted or improper.

The seemingly improper demand might even lead to a boomerang effect. Berkowitz (1969) demonstrated that there were different relationships between the degree of a partner's dependency on the subject and how much work the subject did for him under externally and internally caused dependency. When the partner required help because of circumstances beyond his control, the greater his dependency on the subject, the more work the subject turned out on his behalf. When the partner's own deficiency had made him dependent on the subject, however, increased levels of dependency were associated with a somewhat *lower* level of work in his behalf. Again, it is as if the subjects could succumb to the resentment created by the relatively great demand upon them when this demand was not entirely proper.

Another experiment (Horowitz, 1968) reports direct evidence of the importance of the feeling of behavioral freedom in reactions to "improper" demands. College students were more willing to help a person requiring assistance because of his own inadequacies if they thought they had a freedom of choice than if they believed they were required to help their dependent peer.

The freedom of choice given the subjects in the Horowitz study evidently had lowered their reactance somewhat. In other situations, explicitly telling subjects that they are free either to help someone or not to do so could give them permission to succumb to the reactance they feel and not aid the person in need. According to data reported by Jones (1970), male undergraduates provided with a clear statement of their freedom of choice were most apt to refuse to help a peer who was highly dependent on them, especially if the request implied that the dependent person would have further need of them in the future.

If the pressure to aid another were generally bothersome, we certainly would expect this pressure to be even less tolerable if the individual were in a bad mood. Already feeling bad, demands to help someone would be another source of annoyance, and the pressure might be resisted whether the demands were explicit or only implicit.

Evidence to this effect has been obtained in an experiment by Berkowitz and Connor (1966). College students given a success experience later worked harder for a peer who was dependent on them than did other subjects who had been made to fail on this preliminary task. The "glow of success" evidently enhanced the subjects' motivation to help the individual needing their assistance even though this latter person was not directly connected with the success. Other findings in this study also point to increased resentment as more pressure is felt by the students. Thus, the subjects who had failed earlier and were presumably in a bad mood expressed more dislike for the experiment the greater their partner's dependency on them. The high level of their partner's dependency on them (and the resulting greater felt demand) did not lead to the great expressed hostility in the case of the subjects having the success experience. The earlier failure apparently had lowered the subjects' ability to tolerate the dependent person's implied control over them.[8]

A field experiment by Berkowitz and Macaulay also shows how situationally induced moods can influence a person's willingness to help others (Berkowitz, 1972). Individually interviewed women shoppers were either given praise and approval or only noncommital neutral responses during the course of the questioning. Soon after the completion of the interview a male college student approached each woman and asked her for 40 cents for bus fare, explaining that he had lost his wallet. Other women in the control condition were similarly approached without the preceding interview. The interviewed-but-not-praised shoppers were much much less likely than the control women to give money to the student requesting aid, probably because the time-consuming interview had lowered their willingness to provide this form of help. However, if the women had been praised during the questioning, their motivation to help the student increased; a reliably higher proportion of interviewed-and-praised than interviewed-but-not-praised women gave the money to the boy in need. The interviewer's approval had put the shoppers in a good mood and had enhanced their willingness to be helpful.

More recently, Isen and Levin (1972) demonstrated that this increased helpfulness by people in a good mood is not simply a matter of their being generally compliant. The college men in this study who had a pleasant experience with one person were later more willing than others not having this experience to aid another individual soliciting their help (by agreeing to serve in his experiment) only when their behavior would supposedly benefit rather than distract the other people also participating in the ostensible investigation. Their presumed good feeling only made them more inclined to carry out the helpful activities that were requested of them.

Self-concern also lessens the chances that one person will help another. Several reasons can be advanced for this. For example, as Aronfreed (1970) suggested, the potential help-giver's empathic reactions to the person in need might lead him to give the required assistance. Self-concern conceivably could interfere with this empathy. This individual who thinks only of himself is not likely to experience the anxieties felt by someone in need of help or to anticipate this person's joy at reaching his goal. In addition, of course, a self-concerned person is probably in a fairly bad mood. Whatever the exact cause, for the person who is preoccupied with questions of his self-worth, someone else's call for help, or the obligation to assist someone, is another unwelcome pressure.

This self-concern might arise in a number of ways. Failure and frustration could have this effect, and the Berkowitz-Connor study might perhaps be understood in these terms. Berkowitz has also hypothesized (Berkowitz, 1970) that people might react with this altruism-lessening self-concern when they believe their self-worth is being assessed. An experiment by Brehm and Cole (1966) illustrates this. Male undergraduates were first given either an innocuous introduction to the study that played down its importance or were told that their chances of being a success in life would soon be assessed. This latter statement presumably increased the men's self-concern. Following this introduction, and ostensibly before the experiment proper got under way, in half of the cases another person (the experimenter's confederate) did the subject a favor by giving him a soft drink. When the experiment then supposedly began, the confederate was assigned an arduous task, and the subject was told he could help him if he desired. If the confederate had not done the subject a favor before, about half of the men came to his aid when they had finished their own assignment. On the other hand, 93% of the subjects helped the confederate who had done them a favor earlier but only when the experimental situation was said to be relatively unimportant.

---

[8] More recently, Isen (1970) and Isen and Levin (1972) have replicated and extended this finding.

The previous favor had evidently aroused an obligation to reciprocate, and the men generally complied if they were not self-concerned. When the subjects' chances of success in life had supposedly been measured, however, only about 13% of the men helped the person who had done them a favor earlier. All in all, the men who were induced to be self-concerned evidently resented the pressure they felt to return the confederate's favor, and thus were least likely to help the confederate.

Several experiments by Berkowitz and his students (Berkowitz, 1972) have also shown that heightened self-concern leads to lowered helpfulness. One of these studies also demonstrated that male and female college students typically become concerned about different things. While each subject waited for his or her supposed partner, the experimenter (who was of the same sex as the subject) asked the subject for a favor, clearly indicating this was unrelated to the study: Would he (or she) score some data from another experiment? The subject's output at this time provided a base-line measure of his (or her) typical work rate. After 5 minutes the experimenter said the other person undoubtedly was not going to show up, but added that they would proceed with part of the experiment.

In creating the self-concern variations, the subject was shown pictures of groups, and in two of the three conditions was asked to identify the leader of each group and the group attitude toward this person. The pictures were described as a test of "social sensitivity" in one of these conditions, and as a measure of the individual's "supervisory ability" in the other treatment. It was expected that the women would become especially concerned about their performance on the supposed test of social sensitivity and that the men would be most concerned about doing well on the supervisory ability test. Finally, in a third, control condition, the subjects were only to indicate their liking for the pictures. After the subjects had written their judgments, the experimenter collected them and said he would go off to score these data right away. Each subject was asked to wait until the experimenter returned with the scores. But while waiting, he was asked to help the experimenter again by scoring the other data he had collected from the other study.

Analysis of the change in work rate during this period from the earlier, base-line rate yielded the expected outcome. The men *increased* their work rate after taking either the "social sensitivity" or "preference" tests, but actually slowed down when they thought the test had measured their "supervisory ability." In the case of the women, on the other hand, the *lowest* increase in work arose after they had taken a test of their "social sensitivity." All in all, the men and women seemed to have somewhat different concerns, but whatever made them self-preoccupied, the arousal of self-concern evidently lowered their willingness to aid the person requesting help.

Conclusion

Assuming that people do resent the obligation they feel to aid those who are dependent upon them, then, the next question becomes, how can they be induced to live up to this obligation and be more helpful? One answer, of course, is to minimize those situational conditions that appear to lessen their tolerance for reactance. But it may be that more positive steps can be taken as well. These procedures would (*a*) clearly established helpfulness as socially desirable conduct and (*b*) provide practice in behaving helpfully.

Although conclusive evidence of the part played by social ideals is still lacking, there is some reason to think that normative considerations affect the individual's readiness to come to the aid of persons in need (cf. Berkowitz & Daniels, 1964; Rawlings, 1970; Schwartz, 1970b). Thus, according to London (1970), a number of Germans who had risked their lives to assist Jews seeking to hide from the Nazis had a "strong identification with a moralizing parent" and firm moral convictions themselves.

There is somewhat more direct evidence in a recent replication of the Darley and Latané diffusion-of-responsibility experiment carried out by Horowitz (1971). Using members of either socially oriented groups (fraternities) or campus community service groups as subjects, each person was told either that he and one other subject were alone in the communication network or that there were three other bystanders as well. When the "other subject" had the supposed seizure, the service-group members were more likely to intervene than were the social-group members, especially if the experimental instructions had deliberately reminded them of their group membership. Furthermore, although the social-group subjects displayed the diffusion of responsibility effect noted by Darley and Latané, the presence of the additional bystanders actually served to increase the service-group members' helpfulness. These latter persons had presumably become more aware of their group-shared ideals prescribing helpfulness when they were in the presence of others, particularly when they had been reminded of their group membership, and the more salient ideal might have induced the greater helpfulness.

However, normative considerations alone are often insufficient; the pressure to do the "right" thing may be resented and then avoided. Staub (1971) suggests that children in our society are more apt to be taught prohibitions than rules positively calling for prosocial behavior, so the latter ideals may be relatively weak in most people. Still, if the individual practices adherence to the moral standard, he is all the more likely to act in conformity to it at other times (Rosenhan & White, 1967). In keeping with this possibility Whiting and Whiting (1970) reported that those cultures in which children tend to be most helpful have typically assigned important familial tasks to these youngsters in the course of their socialization. Repetition of socially desirable behavior might well lessen the person's resistance to act in this manner, especially when the conduct is rewarded. All in all, encouraging people to "do their own thing" and to become preoccupied with their own concerns is all too apt to promote selfishness rather than helpfulness (Berkowitz, 1970). If we want greater helpfulness, we probably have to teach people that this is desirable behavior, have them rehearse this action, and reward them when they do assist others.

REFERENCES

Aronfreed, J. The socialization of altruistic and sympathetic behavior: Some theoretical and experimental analyses. In J. Macaulay & L. Berkowitz (Eds.), *Altruism and helping behavior.* New York: Academic Press, 1970.

Berkowitz, L. Resistance to improper dependency relationships. *Journal of Experimental Social Psychology,* 1969, **5**, 283–294.

Berkowitz, L. The self, selfishness and altruism. In J. Macaulay & L. Berkowitz (Eds.), *Altruism and helping behavior.* New York: Academic Press, 1970.

Berkowitz, L. Social norms, feelings, and other factors affecting helping and altruism. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. 6. New York: Academic Press, 1972.

Berkowitz, L., & Connor, W. H. Success, failure and social responsibility. *Journal of Personality and Social Psychology,* 1966, **4**, 664–669.

Berkowitz, L., & Daniels, L. R. Affecting the salience of the social responsibility norm: Effects of past help on the response to dependency relationships. *Journal of Abnormal and Social Psychology,* 1964, **68**, 275–281.

Berkowitz, L., Klanderman, S. B., & Harris, R. Effects of experimenter awareness and sex of subject and experimenter on reactions to dependency relationships. *Sociometry,* 1964, **27**, 327–337.

Brehm, J. *A theory of psychological reactance.* New York: Academic Press, 1966.

Brehm, J. W., & Cole, A. H. Effect of a favor which reduces freedom. *Journal of Personality and Social Psychology,* 1966, **3**, 420–426.

Darley, J. M., & Latané, B. Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology,* 1968, **8**, 377–383.

Horowitz, I. A. Effect of choice and locus of depence on helping behavior. *Journal of Personality and Social Psychology,* 1968, **8**, 373–376.

Horowitz, I. A. The effect of group norms on bystander intervention. *Journal of Social Psychology,* 1971, **83**, 265–273.

Isen, A. M. Success, failure, attention and reactions to others: The warm glow of success. *Journal of Personality and Social Psychology,* 1970, **15**, 294–301.

Isen, A. M., & Levin, P. F. Effect of feeling good on helping: Cookies and kindness. *Journal of Personality and Social Psychology,* 1972, **21**, 384–388.

Jones, R. A. Volunteering to help: The effects of choice dependence, and anticipated dependence. *Journal of Personality and Social Psychology,* 1970, **14**, 121–129.

Latané, B., & Darley, J. M. *The unresponsive bystander: Why doesn't he help?* New York: Appleton-Century-Crofts, 1970.

London, P. The rescuers: Motivational hypotheses about Christians who saved Jews from the Nazis. In J. Macaulay & L. Berkowitz (Eds.), *Altruism and helping behavior.* New York: Academic Press, 1970.

Moscovici, S., & Neve, P. Studies in social influence: I. Those absent are in the right. *European Journal of Social Psychology,* 1971, **1**, 201–214.

RAWLINGS, E. I. Reactive guilt and anticipatory guilt in altruistic behavior. In J. Macaulay & L. Berkowitz (Eds.), *Altruism and helping behavior.* New York: Academic Press, 1970.

ROSENHAN, D. L., & WHITE, G. M. Observation and rehearsal as determinants of prosocial behavior. *Journal of Personality and Social Psychology,* 1967, **5**, 423–431.

ROSENTHAL, A. M. *Thirty-eight witnesses.* New York: McGraw-Hill, 1964.

SCHOPLER, J., & BATESON, N. The power of dependence. *Journal of Personality and Social Psychology,* 1965, **2**, 247–254.

SCHOPLER, J., & MATTHEWS, M. W. The influence of the perceived causal locus of partner's dependence on the use of interpersonal power. *Journal of Personality and Social Psychology,* 1965, **2**, 609–612.

SCHWARTZ, S. H. Elicitation of moral obligation and self-sacrificing behavior: An experimental study of volunteering to be bone marrow donor. *Journal of Personality and Social Psychology,* 1970, **15**, 283–293 (a)

SCHWARTZ, S. H. Moral decision making and behavior. In J. Macaulay & L. Berkowitz (Eds.), *Altruism and helping behavior.* New York: Academic Press, 1970.

SCHWARTZ, S. H., & CLAUSEN, G. T. Responsibility, norms, and helping in an emergency. *Journal of Personality and Social Psychology,* 1970, **16**, 299–310.

SEARS, R. R., MACCOBY, E. E., & LEVIN, H. *Patterns of child rearing.* Evanston, Ill.: Row, Peterson, 1957.

STAUB, E. Helping a person in distress: The influence of implicit and explicit "rules" of conduct in children and adults. *Journal of Personality and Social Psychology,* 1971, **17**, 137–144.

WHITING, J. M. W., & WHITING, B. Children in six cultures. Unpublished manuscript, Harvard University, 1969. Cited by E. Staub, A child in distress: The effect of focusing responsibility on children on their attempts to help. *Developmental Psycyhology,* 1970, **2**, 152–153.

(Received May 1, 1972)

Manuscripts Accepted for Publication in the *Psychological Bulletin*

Type IV Errors Revisited. Paul A. Games (Department of Educational Psychology, 201 Social Sciences Building, Pennsylvania State University, University Park, Pennsylvania 16802).

Organizational, Work, and Personal Factors in Employee Turnover and Absenteeism. Lyman W. Porter (Office of the Dean, Graduate School of Business Administration, University of California at Irvine, Irvine, California 92664) and Richard M. Steers.

The Intellectual Consequences of Perinatal Anoxia. Allen W. Gottfried (69-54 197 St., Fresh Meadows, New York 11365).

A Note on the Alleged Nonexistence of Individual Differences in Reminiscence. H. J. Eysenck (Institute of Psychiatry, De Crespigny Park, Denmark Hill, London, SE5 8AF, England).

A Note on Alleged Individual Differences in Reminiscence and Extraversion–Introversion. Edward N. Peters (Department of Pediatrics, Well Baby Research Unit, 1545 Mt. Hope Avenue, Apartment 10, Rochester, New York 14620).

Laughter in Young Children. Mary K. Rothbart (Department of Psychology, University of Oregon, Eugene, Oregon 97403).

Psychological Bulletin
1973, Vol. 79, No. 5, 318–334

IMPLOSIVE THERAPY AND FLOODING PROCEDURES:

A CRITICAL REVIEW [1]

KENNETH P. MORGANSTERN [2]

Pennsylvania State University

After a brief explanation of implosive theory and technique, the case reports and experimental investigations of implosive therapy and flooding procedures with human subjects are critically reviewed. Much of the research has produced confusing and contradictory results and these findings are discussed in terms of the type of response rehearsed in therapy, exposure time, and cognitive factors. However, since a majority of the experimental reports are so badly confounded, interpretation of the literature is exceedingly difficult. It is concluded that there is, at present, no convincing evidence of the effectiveness of implosion or flooding with human subjects nor is there any evidence that the techniques are superior to systematic desensitization. In addition, the theoretical basis of implosive therapy appears to be unsupported. Finally, serious realistic and ethical questions are raised in regard to the desirability of the clinical use of implosive procedures.

The last decade has witnessed an explosion in the behavior modification literature. Of particular importance has been the development of two principle techniques aimed at the reduction or elimination of conditioned fear and avoidance behavior. One of these procedures, systematic desensitization, has seen the publication of hundreds of reports which have proceeded from initial demonstrations of effectiveness (e.g., Wolpe, 1958), through factorial designs (e.g., Paul, 1966), to analyses of treatment components (e.g., Davison, 1968). Finally, several comprehensive and critical reviews (e.g., Bandura, 1969; Lang, 1969; Paul, 1969b, 1969c; Rachman, 1967; Wilson & Davison, 1971) have helped in the integration and evaluation of the empirical data and in the validation of the theory underlying desensitization.

In contrast, experimental investigations of implosive therapy (the other major technique aimed at the reduction of phobic and avoidance behavior) have been noticeably lacking. Slightly more than 25 reports appear to date with the majority of this research conducted by the innovators and proponents of implosion or by doctoral candidates. The studies have shown little evidence of the systematic progress in methodological sophistication that has characterized the development of desensitization research, nor has there been any integrated review published.

This article, then, examines the case reports and experimental investigations of implosive therapy with human subjects. Animal studies have been excluded, except in a few cases where subhuman analogues have had important implications for theoretical validation or clinical practice. Finally, research in flooding, a similar but distinct technique, is also reviewed and evaluated.

IMPLOSIVE THEORY AND TECHNIQUE

Although slow in coming, the theory and the technique of implosive therapy has now been amply outlined in several reports (Hogan, 1968, 1969; Levis, 1967; Stampfl, 1967; Stampfl & Levis, 1966, 1967a, 1967b, 1968, 1969). During the late 1950s, Thomas Stampfl, with the assistance of George Golias, Robert Hogan, and, later, Donald Levis, developed a learning-theory-based approach to treatment incorporating many of the formulations consistent with psychoanalytic theory. Based upon an extension of Mowrer's two-factor model of learning (e.g., Mowrer, 1939,

---

[1] This article was written while the author was supported in part by a United States Public Health Service Fellowship. Acknowledgment is extended to W. Edward Craighead for his helpful suggestions and comments on an earlier draft of this manuscript.

[2] Requests for reprints should be sent to Kenneth P. Morganstern, Department of Psychology, Pennsylvania State University, 417 Psychology Building, University Park, Pennsylvania 16802.

1947), phobic or "symptomatic" behaviors are seen, according to Stampfl, as end products in a long chain of higher-order conditioning. These behaviors are viewed as attempts to reduce the anxiety associated with the more threatening thoughts and imagery that are supposed to have accompanied the original aversive conditioning. Defensive maneuvers of the patient are seen, then, as a means of avoiding the dangerous associations which elicit anxiety, even if the patient is unaware of them.

The most basic associations are related to sex, rejection, oral and anal functions, aggression, bodily injury, guilt, punishment, and loss of impulse control. Since the latter are more closely associated with the original aversive conditioning events, they are considered to be more important and are more likely to elicit the anxiety response (even though they may be "repressed"), than are the observable "symptoms."

Important in the development of the therapeutic technique was the experimental evidence of Miller (1951), Solomon, Kamin, and Wynne (1953), and Levis (1967) who demonstrated that extinction occurred more rapidly and more fully the closer the extinction situation approximated the original learning conditions. In addition, it was suggested that the more intense the fear evoked, the greater the anxiety decrement. Although Stampfl and Levis (1967a) cite additional evidence supporting this hypothesis, they go well beyond the experimental data (entirely subhuman) in extrapolating the principles to therapy. For example, Stampfl (1967) hypothesizes that psychodynamic themes and conflicts must be present in the original conditioning situations and that these "stimuli" have to be included in therapy. The goal, then, of the implosive therapy therapist is

> to reproduce in the absence of physical pain, as good an approximation as possible the sights, sounds, and tactual experiences originally present in the primary conditioning event. The therapist attempts to achieve this goal by verbal instructions to the patient to imagine the various aversive stimuli hypothesized to be present in the original conditioning events. It is assumed that all aversive cues experienced in imagery function as secondary conditioned stimuli and thus are extinguishable. Exposure to cues representing or approximating these events is likely to elicit relatively high levels of anxiety. Their repetition in the absence of a primary reinforcer, however, should lead to a decrement in their aversiveness through the principle of generalization of extinction and thereby to a reduction in the avoidance reactions which constitute symptomatic behavior . . . [Stampfl & Levis, 1968, p. 33].

The procedure of implosive therapy begins with an initial assessment of the crucial stimuli that are associated with the patient's anxiety. From this assessment an Avoidance Serial Cue Hierarchy is constructed, with conditioned stimuli ordered according to the extent to which they are associated with the original primary reinforcement. Symptom-contingent cues, involving situational and environmental stimuli that are associated with the actual avoidance behavior, are low on the Avoidance Serial Cue Hierarchy and are the ones presented and extinguished first. Hypothesized sequential cues (assumed to be repressed) are higher on the Avoidance Serial Cue Hierarchy and involve dynamic themes such as aggression, punishment, and sexual material. The latter cues are also seen to be related to the patient's symptomatic behavior, but because they are more threatening and more highly avoided, they are extinguished last. According to Stampfl and Levis (1967a), it is not important that the patient accept the dynamic interpretation, the emphasis being simply on the extinction of all the anxiety-evoking cues which provide both motivational and reinforcing properties for perpetuating the patient's avoidance responses.

During implosive therapy, the patient is instructed to "lose himself" in the part that he is playing and "live" the scenes that the therapist describes with genuine emotion and affect. At each stage of the process an attempt is made by the therapist to attain a maximum level of anxiety evocation from the patient by embellishing the scenes with the most horrifying and unrealistic images. When a high level of anxiety is achieved, the patient is held at this level until he implodes—a spontaneous reduction in the anxiety-inducing value of the cues (Stampfl & Levis, 1967a). This procedure is continued until a significant diminution in anxiety has resulted, the therapist gradually ascending the Avoidance Serial Cue

Hierarchy until the anxiety elicited by all stimuli has been extinguished.

LITERATURE REVIEW

Before consideration of the appropriate research, it is important to make a distinction between flooding procedures and the techniques usually employed by implosive therapists. Many reports have confused the two, often using the terms interchangeably. There seems to be quite a difference, however, between flooding in which intense stimuli are simply presented for an extended period of time, and implosive procedures that expose patients to unrealistic, but horrifying and vivid scenes related to the phobic stimuli. As Bandura (1969) has pointed out, "there is considerable difference between exposing people repeatedly to a fearsome collection of rodents without any adverse effects and depicting them eating human flesh [p. 404]." Although research in both areas is presented in this review, it is separately evaluated (whenever possible).

Case Histories

A few researchers (e.g., Hogan [1963],[3] 1966; Levis, 1967; Stampfl & Levis, 1967a, 1968) have reported high degrees of success using implosive therapy over a wide variety of psychopathological problems. These disorders have included "both psychoneurotic disorders such as anxiety, phobic, obsessive, compulsive, and depressive reactions, and psychotic disorders such as affective, schizophrenic, and paranoid reactions [Stampfl & Levis, 1967a, p. 502]."

Hogan (1968) reported an exceptional case in which a severely disturbed woman received eight implosive therapy treatment sessions. The scenes that the woman was asked to visualize during therapy included extreme aggression, punishment, destruction, and rejection, as well as many phobic stimuli which included rats, bugs, snakes, and mice. The treatment, according to Hogan, produced a remarkable cure although no data is presented. However, the author reports that, in addition to shifts on several Minnesota Multiphasic Personality Inventory (MMPI) scales in the direction away from pathology, the patient lost weight, was no longer depressed, showed better contact with reality, and was no longer plagued by guilt, paranoid thoughts, or hypochondriacal symptoms. Finally, fears of death, people, sex, assault, and insects were completely eliminated after treatment.

Baum and Poser (1971) have recently reported two successful cases of flooding procedures in vivo. In the first instance, a woman, described as a severe obsessive-compulsive, was afraid of contracting "cancer germs" from a wide variety of objects, particularly raw meat. One session, in which the patient approached, picked up, and played with raw pork, successfully extinguished her fears. In the second case, a 19-year-old schizophrenic boy, whose main fears were being with girls his own age and public speaking (because of an obsession about "bad breath"), was also successfully treated in one session. The treatment consisted of playing a game with several young women in which he was forced to be very close to them for an extended period of time.

Even Wolpe (1969) reported two cases in which he used flooding procedures. One of the patients was markedly improved after a single session, but in the second case where flooding procedures were employed in vivo, anxiety was noticeably intensified, resulting in hospitalization of the patient. Wolpe concluded that individual differences determine whether a person will be amenable to flooding methods, or will be harmed by them, and emphasized that the risks involved were not justfied, except as a final resort.

Although it is unreasonable to expect many therapists to report failures, Hogan (1966), in fact, reported an unsuccessful attempt by Golias and Hogan in 1961. In this instance, over 50 sessions of implosive therapy with a 29-year-old chronic schizophrenic served only to produce "sociopathic behavior" in the patient, from which "he never recovered."

While several investigators have reported overwhelming success with patients using implosive or flooding techniques, some have reported failures, making Stampfl's early claim of 100% success unjustified. None of these reports have included data, although a few point to "improved MMPI profiles." While interesting, perhaps, case histories have little, if any, formal scientific value, and at best provide very crude hypotheses about cause–effect relationships. All of the reports contain, what Paul (1969a) and Underwood (1957) have termed, "lethal errors" (i.e., a complete confounding between classes of variables) and, therefore, do not allow for any conclusions about the effectiveness of implosion or flooding.

---

[3] R. A. Hogan. The implosive technique: A process of reeducation through the application of principles of learning for emotionally disturbed individuals. Unpublished doctoral dissertation, Western Reserve University, 1963.

Experimental Research: Implosion

A series of research investigations by Hogan and Kirchner offered the first and most optimistic reports on the effectiveness of implosive therapy. Hogan (1966) conducted the initial study of implosion in the early 1960s at the Hawthornden State Hospital in Macedonia, Ohio. Fifty psychotic patients were assigned to either implosive or nonimplosive therapists by a ward physician who was reportedly unaware that a study was being conducted. The groups were equated for sex, age, education, IQ, prior hospitalization, and length of institutionalization. All subjects participated, in addition to the treatments under study, in group therapy, recreation, work, educational programs, and occupational therapy. Results indicated no significant MMPI changes (the only dependent measure used) for subjects who were assigned to nonimplosive therapists, whereas subjects receiving implosive therapy showed significant shifts away from pathology on five MMPI scales. As a further index of success, 18 of the 26 imploded subjects had been discharged from the hospital at a 1-year follow-up, while only eight of 24 control subjects had left.

Kirchner and Hogan (1966) randomly assigned 40 rat-phobic college females to either an implosion or control condition after they were unable to pick up a rat on a pretest. All subjects received treatment together by listening through individual earphones to a tape recording in a language laboratory. Experimental subjects listened to an implosive procedure that described rats biting and attacking them in packs, while the controls were instructed to imagine pleasant scenes while they listened to music. Results indicated that 62% of imploded subjects and 26% of controls handled the rat on the posttest, a statistically significant difference.

In a similar experiment, Hogan and Kirchner (1967) individually treated implosive therapy subjects in one session, the length of treatment varying from individual to individual (averaging 39 minutes). Control subjects, also seen one time individually (average treatment time = 30 minutes), were asked to visualize neutral imagery such as a walk or a relaxed vacation. On a behavioral posttest 67% of imploded subjects and 10% of controls were able to pick up a rat, a statistically significant chi square.

In the final study, Hogan and Kirchner (1968) randomly assigned 30 snake-phobic coeds to either an implosive, eclectic verbal therapy, or bibliotherapy condition. Implosion subjects, seen individually by the senior author for 45 minutes, were asked to imagine snakes bitting them and wrapping so tightly around their necks that they slowly strangled. Subjects in the eclectic verbal therapy group, all treated by the junior author, discussed their interpersonal relationships and their prior experiences with snakes, being reassured by the experimenter that snakes were harmless. Finally, bibliotherapy subjects were handed a book about snakes and their habits, that clarified many of the myths and untruths about snakes. After one treatment session, 70% of the implosive therapy subjects passed the posttreatment behavioral test in contrast to only 40% of eclectic verbal therapy and 10% of bibliotherapy subjects who picked up the snake. Although the difference between implosive therapy and eclectic verbal therapy subjects was not significant, both of these groups were reliably different than bibliotherapy subjects. Following the posttest, the failures of both "control" groups were given one implosive therapy session. Results of this additional session indicated that 67% of these subjects then passed the approach test (although initial treatment and implosive therapy were obviously confounded).

Taken together, the Hogan and Kirchner series offer initially impressive and encouraging findings. On closer inspection, however, it is apparent that the investigations are so replete with methodological deficiencies that it is impossible to draw any firm conclusions about the effectiveness of implosive therapy. One of the major shortcomings is the absence of true control groups in most of the reports. For example, "control" subjects reported by Hogan (1966) received treatment indistinguishable from ordinary hospital routine. Similarly, listening to music (Kirchner & Hogan, 1966) or imagining a relaxed vacation (Hogan & Kirchner, 1967) were rather transparent "pseudotherapies" no matter what rationale was given. Thus the possible placebo effects of a new and radically different treatment were left essentially uncontrolled in these three designs.

Also noteworthy is the apparent disregard for the effects of experimenter bias in the majority of these reports. It is difficult to see, for example, how a ward physician can equate treatment groups on several variables (Hogan, 1966) and still be "unaware that a study was being conducted." Likewise, experimental subjects in two studies (Hogan & Kirchner, 1967, 1968) were observed on the posttest by the same experimenter who administered the treatment condition. Not only is bias obvious, but it would also seem that experimental subjects must have had extreme "demands" created to perform on the behavioral test, since their "therapist" was conducting the assessment. Hogan and Kirchner (1968) dismiss this point as unimportant, citing a previous study (Kirchner & Hogan, 1966) that achieved positive outcomes while controlling for experimenter bias and demand. The two studies, however, are not really comparable. Rather than simply listening to a tape (as in the earlier study), Hogan and Kirchner's (1967, 1968) implosive therapy subjects spent an entire therapy session with the same investigator who then observed their posttreatment performance.

Additional problems in several of the studies also add to the confusion in interpreting the results. Hogan (1966), for example, failed to equate the length of treatment, therapist effects, and pre-MMPI scores in his experimental conditions. Hogan and Kirchner's (1967) implosion subjects received approximately 25% more "treatment" than controls, and the implosive therapy and control subjects treated by Hogan and Kirchner (1968) were seen by different therapists. Since there is a total confounding of experimental condition and a host of other important variables in these three reports, one must seriously question the validity of such control groups.

A number of additional criticisms of experimental design also deserve mention. The use of the MMPI (Hogan, 1966) as the sole dependent measure is questionable as a sign of pathology or of treatment change. Also, it appears likely that most of the subjects in at least one of the studies (Hogan & Kirchner, 1968) were not very fearful to begin with, since a short verbal discussion (eclectic verbal therapy) enabled 40% of the "snake-phobics" in this group to handle the feared object. Finally, in view of the fact that the researchers have been closely associated with Stampfl and Levis (the major proponents of implosive therapy), it is interesting to note the absence of any psychodynamic themes in their treatment procedures, even though they report great success. Thus it appears that at least one of Stampfl's major assumptions has been called into question (Rachman, 1969) by one of the investigators who was instrumental in the development of the theory (Hogan).

Levis and Carrera (1967) randomly assigned 40 patients to an implosive therapy group or to one of three control conditions. The subjects in the implosive therapy group each received one or two 1-hour standard diagnostic interviews, followed by 10 1-hour sessions of implosion. One implosive therapy therapist, experienced in the technique, treated 7 of the 10 subjects in the group, while another experimenter inexperienced with implosion, treated the remaining 3 subjects. Two control groups received "conventional" therapy (insight and supportive) for either 10 hours (CT_1, the same length of treatment as implosive therapy), or for a much longer period (CT_2, mean length = 37 hours). The last control group, no treatment (WT), remained on a waiting list, participating only in pre- and postassessments. Results indicated that imploded subjects showed a significantly greater decrease on 9 out of 10 standard scales of the MMPI (the only dependent variable employed in the study). Further, the authors point out, 35% of pretest scores above

70 (the cutoff point for pathology) for implosive therapy subjects dropped into the normal range after treatment, while control subjects in CT_1, CT_2, and WT decreased only 12%, 9%, and 7%, respectively. Finally, 58% of the scores above 80 dropped for the implosion group (although not necessarily into the "normal" range), whereas no such changes were observed for the controls.

While the investigation was referred to as only "exploratory," several procedural errors in the design prevent any but the most tentative conclusions from being made. The failure to equate groups for length of treatment, cost of treatment (some subjects were assessed fees, others were not), and experience of therapists was evident. Also, different therapists were used for different treatments and the groups were seen, apparently, at different times, thus completely confounding the major independent variable. Finally, even the sole criterion, the MMPI, is confounded since inspection of the data reveals unequated pretest scores for experimental and control groups. The authors report greater decreases on 9 out of 10 MMPI scales for implosive therapy subjects. However, on 7 of these 9 scales, imploded subjects had greater pretest scores, making a statistical explanation (regression toward the mean) as likely as any other for the observed changes. Further, there were no significant differences in posttreatment scores between groups, and when implosive therapy subjects are compared to all controls combined, only two scales show significant pre–post decreases. One of these scales, *Mf*, has little value in pathology assessment, while the other, *Sc*, is difficult to fit into an implosion model of behavior change. The only change that seems predictable from implosive theory would be on the anxiety scale and on this measure one of the control groups showed a *larger* reduction than the implosive therapy group.

Hodgson and Rachman (1970) tested the hypothesis that improvement following implosive therapy was due to a general rapid extinction process which is facilitated by subjecting subjects to preparatory emotional arousal, even if this arousal is unspecific and unrelated to the phobic stimuli. Fifty females, equated on an initial snake approach test, were assigned to one of five treatment groups. Each treatment consisted of a 40-minute tape, followed immediately by an approach test, Eysenck's Psychoticism-Extroversion-Neuroticism inventory, and a fear thermometer. Group 1 listened to an implosion tape, closely following the procedure of Hogan (1966), while Group 2 received 30 minutes of "horror images" followed by 10 minutes of snake images. Group 3 received the same treatment as Group 2, with the exception that the snake images were presented 24 hours later. Group 4 listened to 40 minutes of horror images, and finally, Group 5 listened to 40 minutes of pleasant imagery (kittens instead of snakes).

The only significant findings in the above investigation revealed that improvement on the fear thermometer was limited to control subjects, while implosive therapy subjects reported a significant increase in fear from pre- to posttest. In addition, Group 2 (snake images immediately following horror images) showed a significant increase in fear. These results are in marked contrast to the dramatic improvements reported by Kirchner and Hogan (1966) even though the implosion groups in both studies received very similar treatment (although the authors admit that the expectancies for subjects were not the same). The investigators concluded that the improvement of Group 2, but not Group 3, strengthens the hypothesis that nonspecific emotional arousal "desensitizes" phobic stimuli only if the latter are immediately presented, but in the absence of any significant difference between groups on the avoidance test, a more stringent criterion, it is difficult to make any firm conclusions. A more recent investigation, however, lends some support to the notions of Hodgson and Rachman. In this study, Watson and Marks (1971) reported that imagining either relevant or irrelevant fear cues produced significant improvement on clinical and attitudinal measures, but only irrelevant fear sessions produced significant physiological improvements. Of greater importance was the finding that the imagination of irrelevant items significantly reduced subjective anxiety more than relevant fear sessions.

Component Analysis of Implosion

Several reports of implosive therapy have attempted to analyze the treatment package and break down the procedure into its essential components. Such studies can potentially reveal more about the treatment and improve application, as well as test the theory itself. In addition, such component studies shed additional light, perhaps indirectly, on the effectiveness of implosion treatment.

Kotila[4] investigated the effects on implosion of varying the mode of presentation of the phobic stimuli (continuous versus distributed) and the duration of the session (20 versus 40 minutes). Fifty snake-phobics were individually administered in a single session, one of five treatment conditions, with 10 subjects randomly assigned to each group. Four implosive therapy conditions received either continuous or distributed implosion, and for either 20 or 40 minutes. The fifth condition, Education Therapy entailed giving the subject information about snakes for 6 minutes. Results showed that all treatments reduced fear behavior toward snakes using behavioral and subjective measures at pre-, post-, and follow-up testing. In addition, it was found that the 40-minute continuous implosion group was more successful than the other three implosive therapy conditions. However, the education therapy group, which had been "treated" for only 6 minutes, showed greater fear reduction than any of the implosive therapy methods. Moreover, implosive therapy groups manifested a temporary increase in fear after therapy toward certain previously neutral stimuli, an effect absent in the education therapy group. Although this study involves a few problems (e.g., the possibility of experimenter bias in the assessment procedures and the absence of a no-treatment control), the author's admonition for caution in using implosive therapy seems well taken.

Fazio (1970), in two experiments, attempted to evaluate the separate effects of some of the treatment components of implosive therapy. In Experiment I, 18 "roach-phobics" were randomly assigned to one of three treatment conditions: (*a*) implosive therapy, in which subjects listened to a tape patterned after Kirchner and Hogan (1966); (*b*) "pseudotherapy," which involved visualizing scenes of irrelevant objects, followed by a tape of a frank and rational discussion of irrelevant phobic stimuli (which controlled for the novelty of imagining frightening scenes); and (*c*) a combination of Treatments *a* and *b*. All subjects participated in three 29-minute sessions, each held at 2-week intervals. The results indicated that implosive therapy was not effective in producing any significant reduction in phobic behavior. Indeed, both posttest and follow-up data showed that reduction in fear on most of the behavioral tests was *inversely* related to the amount of implosive therapy experienced by treatment groups, that is, Group 1 showed more fear than Group 3 which, in turn, showed more fear than Group 2.

In the second experiment by Fazio (1970), 29 subjects were randomly assigned to the same three treatment conditions of Experiment I. After two pretests in which each subject attempted to approach a cockroach, three treatment sessions were conducted at 1-week intervals, using the identical tapes as employed in the first study. After completing this phase, each subject then received two additional sessions of a different treatment. Results again showed that implosive therapy subjects failed to show any significant improvement. In fact, three imploded subjects who were not considered phobic on the basis of their pretests, became fearful of the stimulus after only two implosive therapy sessions, leading the author to conclude that the discussion component was the most important factor in producing any improvement.

Prochaska (1971) tested one of the main hypotheses of implosive therapy by presenting test-anxious college students with "symptom cues" (e.g., studying for an exam) or dynamic cues (e.g., parental rejection or castration). Results showed that the symptom cue group indicated significantly more improvement on an intelligence test (the major dependent variable) than any of the controls, while the dynamic cue group differed reliably from only a no-treatment condition, the symptom and dynamic cue groups not differing from each other. In addition, imploded subjects were reported to experience less anxiety on their final exams following treatment and showed modest increases in their grades. The study, however, contains several problems in design, the most important being the complete confounding of experimental conditions with a variety of extraneous uncontrolled variables. For example, some of the subjects were tested in a large group, some in small groups, and some individually. Likewise, treatment conditions varied, with some subjects listening to a taped session through earphones in a language laboratory, while others sat around in a small group listening together to a tape recorder; some groups were as large as 23, others were as small as 3. Finally, the use of the scores of the intelligence tests as criteria for reduced test anxiety is highly questionable, particularly in view of the fact that the small changes that were reported were accompanied with neither mention of reliability nor standard error of measurement.

---

[4] R. R. Kotila. The effects of education and four varieties of implosive therapy on fear of snakes. Unpublished doctoral dissertation, Washington State University, 1969.

Comparative Studies with Systematic Desensitization: Implosive Research

Both implosive therapy and flooding have been compared experimentally to systematic desensitization in several studies. From such investigations several issues may be resolved, including the comparisons of effectiveness, the length of treatment, and the emotional stress that subjects undergo in therapy. In addition, the very theoretical foundations on which each treatment is based can be subjected to empirical tests, since very different mechanisms purportedly underlie each approach. Systematic desensitization, for example, is often explained in terms of reciprocal inhibition in which deep muscular relaxation reciprocally inhibits (counterconditions) anxiety using a graduated hierarchy, with the result that the subject experiences very little anxiety during treatment. On the other hand, implosive and flooding techniques based on extinction principles, subject the patient to intense anxiety for prolonged periods of time.

Eysenck (1968) and Wolpe (1958; Wolpe & Lazarus, 1966) predicted that subjects who were presented with anxiety-eliciting stimuli that were too intense (e.g., in implosion) would actually become "sensitized" to the stimuli and levels of anxiety and avoidance behavior would increase. Stampfl and Levis, in contrast, would argue that if subjects did not experience intensive emotion, then little extinction would result. It appears, then, that the generally accepted theories underlying each technique cannot adequately account for the successes of the other treatment. Comparative studies, therefore, will not only answer questions of "economy" and efficiency, but will also provide very important data for the different theoretical assumptions.

To date, nine studies have appeared testing the relative efficacy of desensitization and implosion. Three of these investigations have shown systematic desensitization to be superior to implosive therapy (implosion subjects either not differing from controls or actually becoming worse after treatment), with the remaining six showing no significant differences.

A well-designed study was performed by Mealiea ([1967],[5] Mealiea & Nawas, 1971) who randomly assigned 50 snake-phobic coeds to one of five groups (systematic desensitization, implosive therapy, and three controls) after they were matched on a behavior test and a fear thermometer. Results indicated that the systematic desensitization group significantly showed the greatest decrease in snake-phobic behavior, while imploded subjects did not differ from any of the controls. A 1-month follow-up revealed that systematic desensitization subjects continued to exhibit the greatest decrease in phobic behavior, while implosive therapy subjects now showed the least amount of change. The author concluded, then, that the counterconditioning model was well supported and that the previous success of some of the implosive studies were either due to confounding or extraimplosive components, such as reality testing and supportive aspects of the therapist.

---

[5] W. L. Mealiea. The comparative effectiveness of systematic desensitization and implosive therapy in the elimination of snake phobia. Unpublished doctoral dissertation, University of Missouri, 1967.

Using a similar design, nearly identical results were reported by Willis ([1968],[6] Willis & Edwards, 1969) supporting the superiority of desensitization treatment. However, this study is marred by the fact that desensitization and implosive therapy groups were equated not on an actual behavior test, but on the basis of their Fear Survey Schedule scores. In view of a recent study by Bernstein (in press) who found that 96% of subjects classified as "phobic" on the basis of their Fear Survey Schedule score picked up and handled the phobic object on a behavioral test, it is conceivable that the groups in Willis' investigation differed in actual avoidance behavior before treatment began.

Barrett ([1968],[7] 1969) also concluded that systematic desensitization was the preferred choice of treatment. Even though initial findings showed no significant differences between implosion and desensitization on any of the dependent measures used, further analysis of the data revealed subtle differences between the groups. Barrett noted that although the means of both groups were identical, the variances were not. Desensitization subjects, for example, showed a consistent and continuous reduction in variance from preassessment to postassessment to follow-up. Implosion subjects, on the other hand, revealed a significantly greater variance change than both systematic desensitization and controls, showing a significant reduction in variance from pre- to postassessment, but an increase from postassessment to follow-up. Noting the arguments of Bergin (1963, 1966, who emphasized the possibility of "deterioration effects" from therapy which may be hidden when only group means are analyzed), Barrett showed that some implosive therapy subjects improved, but others got worse, whereas desensitization subjects were considerably less variable.

Six comparative studies have reported essentially no significant differences between experimental treatments in producing behavioral change (Borkovec;[8] Carek;[9] Horowitz, 1970; Jacobson;[10] Kirts;[11] McGlynn[12]). However, the investigations by Kirts (who used a conditioned galvanic skin response as the dependent variable in comparing "Wolpe" and "Stampfl" groups) and Horowitz (who manipulated treatment conditions through different hypnotic suggestions) must be considered analogue studies and, therefore, generalization to actual desensitization and implosion treatment may be especially difficult. It is important to note, however, that Horowitz's implosion subjects, while equal to desensitization (relaxation) subjects at posttreatment assessment, were no longer significantly different than controls at a follow-up.

The study by Jacobson reported changes in pulse rates for implosive therapy and desensitization subjects in the predicted directions, and the author concluded that both treatments were supported, although it is difficult to justify this interpretation in view of the fact that there were no significant differences on verbal or motor measures. It does indicate the need, however, to use all three types of measures in assessing anxiety during treatment and at pre-, post-, and follow-up assessment (a suggestion previously emphasized by several researchers, e.g., Craighead, in press; Lang, 1969; Mathews, 1971).

---

[6] R. W. Willis. A study of the comparative effectiveness of systematic desensitization and implosive therapy. Unpublished doctoral dissertation, University of Tennessee, 1968.

[7] C. L. Barrett. Systematic desensitization therapy (SDT) versus implosive therapy (IT): A comparative study of the efficiency of two behavior therapies in reducing snake phobic behavior in otherwise normal adults. Unpublished doctoral dissertation, University of Louisville, 1968.

[8] T. D. Borkovec. The comparative effectiveness of systematic desensitization and implosion therapy and the effect of expectancy manipulation on the elimination of fear. Unpublished doctoral dissertation, University of Illinois, 1970.

[9] R. G. Carek. A comparison of two behavioral therapy techniques in the treatment of rat "phobias." Unpublished doctoral dissertation, University of Iowa, 1969.

[10] H. A. Jacobson. Reciprocal inhibition and implosive therapy: A comparative study of a fear of snakes. Unpublished doctoral dissertation, Memphis State University, 1970.

[11] S. M. L. Kirts. An experimental analogue comparing two models of psychotherapy: Reciprocal inhibition and implosive therapy. Unpublished doctoral dissertation, University of Southern California, 1968.

[12] F. D. McGlynn. Systematic desensitization, implosive therapy and the aversiveness of imaginal hierarchy items. Unpublished doctoral dissertation, University of Missouri, 1968.

In contrast to many of the studies reported above which have often been marred by various methodological deficiencies, the study by Borkovec (see Footnote 7; 1972) represents a sophisticated and well-controlled design. On the basis of a stringent screening process, 50 undergraduate females were randomly assigned from stratified blocks based on their approach scores to one of eight treatments. The treatment conditions consisted of four therapy groups (systematic desensitization, implosive therapy, avoidance response placebo, and a no-therapy control) factorially crossed with two expectancy conditions (positive or no expectancy). Of particular interest was the "avoidance-placebo" control which tested whether subjects who were instructed to actually make avoidance responses would, in fact, show increases in their phobic behaviors since both desensitization and implosive therapy models assert that avoidance behaviors reinforced by anxiety reduction are responsible for the maintenance of conditioned anxiety.

All experimental subjects were seen for four 1-hour sessions at 1-week intervals. The results indicated that both desensitization and implosion (regardless of expectancy condition) produced significantly lower physiological arousal (pulse rates) and tended to produce decreases in self-reported fear, although the two treatments did not significantly differ from each other. The fact, however, that the avoidance response group also showed nearly equal improvement raises a possible problem for both systematic desensitization and implosive therapy theory. Borkovec (1972) adopted an habituation model (Lader & Mathews, 1968) to explain these findings, suggesting that the important aspect of therapy may be the repetition of conditioned stimuli in the absence of any primary aversive consequences (the one component common to systematic desensitization, implosive therapy, and the avoidance control groups).

Of greater interest in the study is the finding of a strong expectancy effect on overt behavior measures of fear and, to a lesser degree, on subjects' verbal reports. Even though desensitization and implosion subjects showed decreases in physiological arousal, they did not differ from no-treatment controls on the avoidance test if they did not have the expectancy that they would, in fact, change.

Borkovec concluded that traditional learning mechanisms (e.g., counterconditioning or extinction) may explain physiological changes in therapy, but it is the cognitive factors (e.g., expectancy or demand characteristics) that produce subjective and overt motor differences. These findings are consistent with those of Dee[13] who found that positive instructions preceding implosive therapy had a weak but facilitating effect in reducing fear. The results, however, stand in contrast to the findings of Layne[14] who found no significant expectancy effect in a study in which implosion, listening to music, and no treatment produced the same amount of behavioral change.

Flooding

The term "flooding" was coined by Polin in 1959, although there were many previous suggestions that extinction could be successfully accomplished by arranging intense emotional reactions to take place in the presence of phobic stimuli (Holland & Skinner, 1961; Masserman, 1943; Solomon et al., 1953; Solomon & Wynne, 1954). The first implementation of the technique with human subjects was reported by Malleson (1959) who employed a hierarchy similar to that used in desensitization, although instead of encouraging relaxation, Malleson instructed the client to feel more frightened as he proceeded through the items. The results showed that the client, after an initial increase in distress, showed rapid recovery and a complete cure. Essentially, the same promising results were reported in another early study by Boulougouris and Marks (1962) who found complete cures in a number of different phobic patients after flooding sessions.

Wolpin and Raines (1966) treated six snake-phobic women for a maximum of five sessions using one of three experimental conditions: (*a*) visualization of a 20-item hierarchy with no training or instructions of any kind regarding relaxation; (*b*) completion of a 20-step hierarchy with muscles deliberately tense; and (*c*) presentation of the most intense scenes from the top of the hierarchy with no training or instructions in regard to either relaxation or tension. After four treatment sessions, all subjects were able to handle and fondle the originally feared snake and at a follow-up, 3 to 5 weeks later, five of the subjects were still able to hold the phobic object. Although the investigators concluded that some form of desensitization or extinction was probably the major factor in therapy, they readily conceded that just as important a factor may be the strong fulfillment-of-demand characteristics. The most serious deficiency of this study is, obviously, the fact that few conclusions can be drawn from an experiment having two subjects in each treatment condition. Additional limitations stem from the failure to ensure equal length of treatment, and the absence of an attention-placebo or no-treatment control.

---

[13] C. K. Dee. Instructions and the extinction of a learned fear in the context of taped implosive therapy. Unpublished doctoral dissertation, University of Iowa, 1970.

[14] C. C. Layne. The effects of suggestion in implosive therapy for fear of rats. Unpublished doctoral dissertation, Southern Illinois University, 1970.

Rachman (1966) used flooding in the treatment of three subjects with a demonstrated aversion to spiders, comparing them to three subjects who had previously undergone desensitization treatment and three no-treatment controls. Flooding subjects were instructed to imagine a given scene for 2 minutes at a time, while the experimenter attempted to "vivify" the experience so that maximum anxiety would be elicited. Treatment lasted for 10 sessions, at 2-week intervals, with each session averaging 10 scene presentations.

Results indicated that flooding and control subjects showed no improvement on either an avoidance test or a fear thermometer, whereas matched desensitization subjects had markedly improved. The use of such a short exposure (2 minutes) must be questioned, however, particularly in view of the fact that most implosive studies have used much longer periods for scene presentations. In fact, one of the crucial factors in implosion and flooding may be the exposure time of visualization. Since so few subjects were seen, and since flooding, control, and desensitization subjects were seen at different times, caution should be taken in drawing firm conclusions from this investigation.

A more recent study (Miller & Levis, 1971) addressed itself specifically to the variable of exposure and is relevant here even though the treatment was termed "implosive therapy." In this investigation, 40 high school snake-phobic girls were assigned to treatment in either a 0-, 15-, 30-, or 45-minute exposure condition. The group receiving a 15-minute exposure displayed the most fears as indexed by contact and distant measures with no differences shown between the other three conditions on any of the measures. The authors note, however, that a significant num- of subjects in each group showed improvement on the approach test. Miller and Levis (1971) interpret their findings in the following way:

> [Subjects given a short exposure, 0 minutes] are capable of defending against complete sensory input of the stimuli associated with the phobic test stimulus (PTS) and/or against the redintegration of additional fear eliciting thoughts or images related to the PTS. With more exposure, these defenses may well break down resulting in an increase in the level of fear. Termination of PTS at this point (15 minutes) would be expected to prevent extinction and result in the conservation of a high level of fear. However with continued exposure, these new cues should undergo an extinction decrement [p. 20].

Although the speculations of Miller and Levis (1971) are theoretically consistent with others (Rachman, 1969; Staub, 1968), it appears that the actual exposure times in their study are very much different than those in previous investigations. For example, Rachman's (1966) failure to show improvement from flooding has been explained in terms of a very short exposure time (2 minutes), while Wolpin and Raines' (1966) findings have been interpreted as successful because of their "long" exposure (10 minutes). It is obvious, however, that the latter's "long" exposure time is much shorter than the duration of Miller and Levis' "short" exposure (15 minutes). The results of each study, therefore, cannot be theoretically explained in terms of relative (short versus long) exposure, since the actual duration of stimulus presentation varies, in fact, from one study to another. It is important, then, if such important theoretical issues are to be resolved, that future research use consistent procedures.

An important series of investigations, briefly mentioned here, are those of Morrie Baum. Using primarily subhuman subjects, Baum has contributed a number of important empirical studies demonstrating the success of flooding (or "response prevention," as the procedure has been termed) in the elimination of avoidance behavior (cf. Baum, 1970, for a review and discussion of this research). Although generalization to human subjects may at times be difficult from such animal studies, the parsimony of Baum is refreshing and it seems quite likely that laboratory analogues such as these will shed considerable light on similar procedures with humans. A few recent investigations (e.g., Heath;[15] Hughes;[16] Poppen[17]) have already attempted comparative studies between flooding and desensitization with subhumans that have been quite suggestive.

Flooding versus Systematic Desensitization

Five comparative studies of systematic desensitization and flooding have been reported, indicating in general, equivocal and contradictory findings, and offering little conclusive evidence for the superiority of either procedure. The first investigation was conducted by Strahley[18] who found a counterphobic treatment (flooding) to be significantly more effective than desensitization in reducing fear on the Fear Survey Schedule and an avoidance test, although there were no significant differences on two additional measures. In view of several problems in design (absence of appropriate no-treatment and "attention-placebo" controls, and procedural inconsistencies both between and within treatment conditions), caution must be taken in drawing any strong inferences. Strahley, in fact, concluded that the success of flooding was probably more a function of extratherapy variables than of an intrinsic superiority of the technique.

In an analogue investigation, Brock[19] found no significant differences between a flooding group and a desensitization treatment in the extinction of avoidance behavior. However, lack of no-treatment controls and unusual treatment procedures make the study difficult to generalize to the clinical use of desensitization and flooding.

De Moor (1970) reported that both desensitization and flooding produced significant reductions in fear, the data also suggesting that desensitization was more effective since flooded subjects returned to their pretest level on follow-up testing. The study, however, can be criticized for several procedural errors. De Moor stated that desensitization subjects received a maximum of 10 minutes of relaxation training (method unspecified) in each session. Other studies (e.g., Paul, 1966, 1969d) have reported an average of 25 minutes as the minimum time necessary to produce complete muscular relaxation. In addition, the fact that systematic desensitization subjects spent up to half their 20-minute sessions becoming relaxed, means that only 10 minutes per session were left for desensitization proper to be conducted. Flooding subjects, on the other hand, received 20 minutes of treatment and even though this was twice the amount of "treatment" that desensitization subjects received, exposure time may have still been too short, in view of recent studies (e.g., Kotila, see Footnote 3; Miller & Levis, 1971).

Calef and MacLean (1970), after treating "speech-anxious" college students, reported that both flooding and desensitization produced significantly more improvement than controls on the Taylor Manifest Anxiety Scale and the Personal Report of Confidence as a Speaker. Although the authors concluded that reaction inhibition (flooding) was the more efficient technique (since it was a simpler procedure), deficiencies in design, measures, and statistical analysis make the study of little value (Blanchard, 1971). For example, the use of the Personal Report of Confidence as a Speaker as an adequate dependent variable of change must be seriously questioned. Blanchard (1971) has pointed out that such self-reports are of value only if they are accompanied by actual behavior measures. This can be seen in a study by Karst and Trexler (1970) who found speech-anxious subjects significantly improved on verbal measures, including the Personal Report of Confidence as a Speaker, but did not differ from controls on behavioral assessment. Other problems include the complete confounding of treatment by group and of treatment by therapist (since only one therapist was used) and the inclusion of both high- and low-fearful subjects in the same treatment group. Finally, the statistical analysis of Calef and MacLean presents so many gross errors, that the reported conclusions are completely unsupported by the data (Blanchard, 1971).

---

[15] G. H. Heath. A comparison of reciprocal inhibition and flooding in decreasing the strength of a conditioned avoidance response. Unpublished doctoral dissertation, Temple University, 1968.

[16] H. H. Hughes. An investigation of the assumptions of implosive therapy and systematic desensitization using an animal analogue. Unpublished doctoral dissertation, University of Georgia, 1970.

[17] R. L. Poppen. Counterconditioning of conditioned suppression. Unpublished doctoral dissertation, Stanford University, 1968.

[18] R. F. Strahley. Systematic desensitization and counterphobic treatment of an irrational fear of snakes. Unpublished doctoral dissertation, University of Tennessee, 1965.

[19] L. D. Brock. The efficacy of various extinction procedures on a conditioned avoidance response in humans: An experimental analogue. Unpublished doctoral dissertation, Southern Illinois University, 1967.

The most recent comparative study (Boulougouris, Marks, & Marset, 1971) offers the only clear evidence that flooding might be superior to desensitization (on at least some behavioral measures). Sixteen phobic patients were seen in a crossover design, in which they received six sessions of flooding followed by six sessions of desensitization, or received the same two treatments in the reverse order. Results showed that both experimental conditions produced significant improvement on clinical ratings, although flooding subjects were significantly better on some of the items. Physiological assessment yielded similar findings and on no measure was desensitization superior to flooding. The departures from typical desensitization and flooding make the results, of course, difficult to relate to clinical practice. In addition, it is interesting to note that several patients in the flooding procedure felt that they were "challenged" to prove they could face the phobic stimulus—a demand not shared by the desensitization subjects.

DISCUSSION

It would appear that many of the empirical reports of implosion and flooding have resulted in contradictory and seemingly paradoxical findings. For example, the results of Rachman (1966) whose subjects showed no improvement from flooding, stand in marked contrast to the positive outcomes of Wolpin and Raines (1966). Wilson (1967) noted that Wolpin and Raines' subjects were asked in their imagination to rehearse accomplishment of the final criterion of fear reduction, while Rachman's subjects were required to imagine scenes so horrible that they might only occur in nightmares. Wilson concluded that the effects of flooding (or implosion) depend on the *type of responses* which are rehearsed in therapy. Staub (1968), however, pointed to a few studies reported by Hogan and Kirchner (1967, 1968) and noted that the latter investigators achieved considerable success and still had subjects imagine extremely frightening scenes which were irrelevant to the final criterion responses. The important variable, Staub suggested, which determines the effectiveness of implosion and flooding, is the duration of continuous exposure to the fear-evoking stimuli.

In a later article, Rachman (1969) also addressed himself to the question of exposure. He noted that premature termination of a treatment session may result in an exacerbation rather than a reduction in anxiety. This possibility was raised by Eysenck (1968) and Wolpe (1958, 1969) and can also be deduced from the maximal habituation theory of Lader (Lader & Mathews, 1968; Lader & Wing, 1966). Rachman (1969) noted that, according to this latter theory, desensitization proceeds more effectively when the patient is at a low level of activation. If anxiety-evoking stimuli are introduced when the patient is at a very high level of arousal, an increase in the magnitude of anxiety may result. Implosion or flooding, then, may be explained by assuming that the subject's level of arousal is eventually reduced (as a consequence of the prolonged and intense stimulation) to the point where he can habituate rapidly to the anxiety-evoking stimuli. In other words, "the subject is brought to a point of rapid habituation by first exhausting his capacity for emotional arousal [Rachman, 1969, p. 299]."

The recent report by Borkovec (1972) also suggested that an habituation model might explain the apparently paradoxical finding that an avoidance-placebo group (which actually rehearsed avoidance responses to phobic stimuli) showed considerable improvement in fear reduction. He noted, however, that while learning mechanisms were able to account for physiological changes (in implosion and desensitization), improvement in subjective and overt behaviors could be best explained in terms of cognitive factors, particularly expectancy.

Several animal studies have shown more consistent success with flooding procedures than investigations with human subjects have, although several problems still remain. Baum (1970) has pointed out, for example, that the two-factor theory cannot explain the fact that animals still show overt signs of fear even though the avoidance response is no longer present. Moreover, Wilson and Davison (1971), in a comprehensive review of the animal literature, concluded that the counterconditioning procedure was more effective in reducing conditioned fear than the orthodox classical extinction procedures. Thus it appears that even in the subhuman research, flooding procedures have resulted in equivocal outcomes.

Combining the findings, then, of both human and subhuman research, it seems possible to propose that prolonged exposure (if repeated long enough) may lower physiological arousal to a degree sufficient for subjects to make approach responses (covert or real). In human subjects, cognitive factors may facilitate this process by creating strong demands to perform or by establishing positive expectancies for change, resulting in a reduction of both avoidance behavior and subjective anxiety. Further, subjects may then be motivated to reinterpret their physiological reactions (as Schachter & Singer, 1962, have suggested) so that they are more consonant with the subject's overt behavior.

It is apparent from this review that experimental investigations of implosive therapy and flooding have failed to demonstrate convincingly the effectiveness of these procedures in the elimination of phobic and avoidance behaviors. Although a few studies have reported positive findings (the majority conducted by staunch advocates of implosion), most of these investigations are so badly marred by methodological deficiencies, misinterpretations of data, invalid assumptions, and lack of proper controls that it is impossible to conclude any cause–effect relationships between treatment and outcome.

An additional difficulty with the majority of the reports is the fact that most of the studies have employed analogue designs in which the target behaviors have been fear and avoidance of snakes, insects, or rats. The problems of generalization from analogue to clinical populations have been previously discussed by several researchers (e.g., Cooper, Furst, & Budger, 1969; Lader, 1967; Lang, 1969) and the same criticisms are applicable to much of the implosion literature. Of particular importance is the finding that the verbal reports of analogue fears are often poorly correlated with actual behavioral measures (e.g., Davison, 1968; Geer, 1965; Lang, 1968; Lang, Lazovik, & Reynolds, 1965). Thus, studies that have used only verbal measures to select subjects are especially difficult to interpret. Moreover, Bernstein (in press) has reported two studies showing that even performance on actual behavioral avoidance tests may be strongly influenced by expectancy, the assessment setting, and the demand characteristics of the testing situation.

It would seem, therefore, that very subtle factors may have greatly influenced many of the investigations reviewed here, and the important implications of Bernstein's findings, like the suggestions of Borkovec (1972), need to be considered fully and systematically examined in future research. The role of cognitive processes in systematic desensitization has recently attracted great interest (e.g., Davison & Valins, 1968, 1970; Gaupp, Stern, & Galbraith, 1972; Wilkins, 1971) and it appears likely that similar factors may, likewise, account for much of the variance in implosive and flooding procedures. It remains, however, an empirical question that has so far (with few exceptions) not been tested.

Conclusions

1. The literature in implosive therapy and flooding has not only been less extensive and less sophisticated than research in desensitization, it has also shown the treatments to be far less effective.

2. The comparative studies of implosion and flooding with systematic desensitization have yielded similarly equivocal results, with only one study suggesting the superiority of flooding procedures.

3. In view of these findings, it seems reasonable to posit the view that in many of the instances in which implosion and flooding have proven successful, it has been the demand characteristics of the situation, the expectancy of the subjects, or other extratherapy variables that have contributed the major portion of the outcome variance.

4. The theoretical basis of implosion also seems, at present, to be unsupported. Certainly the dynamic speculations of Stampfl and Levis seem unwarranted since the technique, when it has worked, has more often than not omitted such psychoanalytic aspects from treatment.

5. The few studies that have shown positive treatment effects indicate the need for more systematic and well-controlled future research, particularly in the investigation of cognitive factors in therapy.

6. Finally, even if the effectiveness of implosion were proven, there would still be realistic and ethical considerations to be made. It is quite likely that very few clinicians would find implosion "palatable" and would, therefore, be unable or unwilling to utilize such a technique. One must seriously question the desirability of having a client undergo such stress and trauma in the name of treatment, when systematic desensitization is at least as effective a procedure.

These factors may very well account for the paucity of research in the area. Nevertheless, they remain academic until the more basic empirical validation and theoretical support are demonstrated. It appears, then, that the early questions of Stampfl and Levis (1967a) are still unresolved and their cautions are still prominent, namely, that "further research on the efficacy of the technique, theory, and interaction of the two is still needed before any definite claims [for implosive therapy] can be made [p. 502]."

References

Bandura, A. *Principles of behavior modification.* New York: Holt, Rinehart & Winston, 1969.

Barrett, C. L. Systematic desensitization versus implosive therapy. *Journal of Abnormal Psychology,* 1969, **74**, 587–592.

Baum, M. Extinction of avoidance responding through response prevention (flooding). *Psychological Bulletin,* 1970, **74**, 276–284.

Baum, M., & Poser, E. G. Comparison of flooding procedures in animals and man. *Behaviour Research and Therapy,* 1971, **9**, 249–254.

Bergin, A. E. The effects of psychotherapy: Negative results revisited. *Journal of Counseling Psychology,* 1963, **10**, 244–250.

Bergin, A. E. Some implications of psychotherapy research for therapeutic practice. *Journal of Abnormal Psychology,* 1966, **71**, 235–246.

Bernstein, D. A. Behavioral fear assessment: Anxiety or artifact. In H. Adams & P. Unikel (Eds.), *Issues and trends in behavior therapy.* Springfield, Illinois: Charles C Thomas, 1973, in press.

Blanchard, E. B. A comparison of reciprocal inhibition and reactive inhibition therapies in the treatment of speech anxiety: A methodological critique. *Behavior Therapy,* 1971, **2**, 103–106.

Borkovec, T. D. Effects of expectancy on the outcome of systematic desensitization and implosive treatments for analogue anxiety. *Behavior Therapy,* 1972, **3**, 29–40.

Boulougouris, J. C., & Marks, I. M. Implosion (flooding): A new treatment for phobias. *British Medical Journal,* 1962, **2**, 721–723.

Boulougouris, J. C., Marks, I. M., & Marset, P. Superiority of flooding (implosion) to desensitization for reducing pathological fear. *Behaviour Research and Therapy,* 1971, **9**, 7–16.

Calef, R. A., & MacLean, G. D. A comparison of reciprocal inhibition and reactive inhibition therapies in the treatment of speech anxiety. *Behavior Therapy,* 1970, **1**, 51–58.

Cooper, A., Furst, J. B., & Budger, W. H. A brief commentary on the usefulness of studying fears of snakes. *Journal of Abnormal Psychology,* 1969, **74**, 413–414.

Craighead, W. E. The role of muscular relaxation in systematic desensitization. In R. Rubin (Ed.), *Advances in behavior therapy.* Vol. 5. New York: Academic Press, 1973, in press.

Davison, G. C. Systematic desensitization as a counterconditioning process. *Journal of Abnormal Psychology,* 1968, **73**, 91–99.

Davison, G. C., & Valins, S. On self-produced and drug-produced relaxation. *Behaviour Research and Therapy,* 1968, **6**, 401–402.

Davison, G. C., & Valins, S. A reply to Wolpe's critique regarding self-produced and drug-produced relaxation. *Behaviour Research and Therapy*, 1970, **8**, 107–108.

De Moor, W. Systematic desensitization versus prolonged high intensity stimulation (flooding). *Journal of Behavior Therapy and Experimental Psychiatry*, 1970, **1**, 45–52.

Eysenck, H. J. A theory of the incubation of anxiety/fear responses. *Behaviour Research and Therapy*, 1968, **6**, 309–322.

Fazio, A. F. Treatment components in implosive therapy. *Journal of Abnormal Psychology*, 1970, **76**, 211–219.

Gaupp, L. A., Stern, R. M., & Galbraith, G. G. False heart-rate feedback and reciprocal inhibition by aversion relief in the treatment of snake avoidance behavior. *Behavior Therapy*, 1972, **3**, 7–20.

Geer, J. H. The development of a scale to measure fear. *Behaviour Research and Therapy*, 1965, **3**, 45–53.

Hodgson, R. J., & Rachman, S. An experimental investigation of the implosion technique. *Behaviour Research and Therapy*, 1970, **8**, 21–28.

Hogan, R. A. Implosive therapy in the short term treatment of psychotics. *Psychotherapy: Theory, Research and Practice*, 1966, **3**, 25–31.

Hogan, R. A. The implosive technique. *Behaviour Research and Therapy*, 1968, **6**, 423–432.

Hogan, R. A. Implosively oriented behavior modification: Therapy considerations. *Behaviour Research and Therapy*, 1969, **7**, 177–184.

Hogan, R. A., & Kirchner, J. H. A preliminary report of the extinction of learned fears via short term implosive therapy. *Journal of Abnormal Psychology*, 1967, **72**, 106–111.

Hogan, R. A., & Kirchner, J. H. Implosive, eclectic verbal and bibliotherapy in the treatment of fears of snakes. *Behaviour Research and Therapy*, 1968, **6**, 167–171.

Holland, J. G., & Skinner, B. F. *The analysis of behavior*. New York: McGraw-Hill, 1961.

Horowitz, S. L. Strategies within hypnosis for reducing phobic behavior. *Journal of Abnormal Psychology*, 1970, **75**, 104–112.

Karst, T. O., & Trexler, L. D. Initial study using fixed-role and rational-emotive therapy in treating public-speaking anxiety. *Journal of Consulting and Clinical Psychology*, 1970, **34**, 360–366.

Kirchner, J. H., & Hogan, R. A. The therapist variable in the implosion of phobias. *Psychotherapy: Theory, Research, and Practice*, 1966, **3**, 102–104.

Lader, M. H. Palmar skin conductance measures in anxiety and phobic states. *Journal of Psychosomatic Research*, 1967, **16**, 271–281.

Lader, M. H., & Mathews, A. M. A physiological model of phobic anxiety and desensitization. *Behaviour Research and Therapy*, 1968, **6**, 411–421.

Lader, M. H., & Wing, L. *Physiological measures, sedative drugs and morbid anxiety*. London: Oxford University Press, 1966.

Lang, P. J. Fear reduction and fear behavior: Problems in treating a construct. In J. M. Shlien (Ed.), *Research in psychotherapy*. Washington, D. C.: American Psychological Association, 1968.

Lang, P. J. The mechanics of desensitization and the laboratory study of human fear. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status*. New York: McGraw-Hill, 1969.

Lang, P. J., Lazovik, A. D., & Reynolds, D. J. Desensitization, suggestibility, and pseudotherapy. *Journal of Abnormal Psychology*, 1965, **70**, 395–402.

Levis, D. J. Implosive therapy, Part II: The subhuman analogue, the strategy, and the technique. In S. G. Armitage (Ed.), *Behavior modification techniques in the treatment of emotional disorders*. Battle Creek, Mich.: V. A. Publication, 1967.

Levis, D. J., & Carrera, R. N. Effects of 10 hours of implosive therapy in the treatment of outpatients: A preliminary report. *Journal of Abnormal Psychology*, 1967, **72**, 504–508 .

Malleson, N. Panic and phobia: A possible method of treatment. *Lancet*, 1959, **1**, 225–227.

Masserman, J. H. *Behavior and neurosis*. Chicago: University of Chicago Press, 1943.

Mathews, A. M. Psychophysiological approaches to the investigation of desensitization and related procedures. *Psychological Bulletin*, 1971, **76**, 73–91.

Mealiea, W. L., & Nawas, M. M. The comparative effectiveness of systematic desensitization and implosive therapy in the treatment of snake phobia. *Journal of Behavior Therapy and Experimental Psychiatry*, 1971, **2**, 85–94.

Miller, B. V., & Levis, D. J. The effects of varying short visual exposure times to a phobic test stimulus on subsequent avoidance behavior. *Behaviour Research and Therapy*, 1971, **9**, 17–21.

Miller, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

Mowrer, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychological Review*, 1939, **46**, 553–565.

Mowrer, O. H. On the dual nature of learning: A reinterpretation of "conditioning" and "problem solving." *Harvard Educational Review*, 1947, **17**, 102–148.

Paul, G. L. *Insight vs. desensitization in psychotherapy: An experiment in anxiety reduction*. Stanford: Stanford University Press, 1966.

Paul, G. L. Behavior modification research: Design and tactics. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status*. New York: McGraw-Hill, 1969. (a)

Paul, G. L. Outcome of systematic desensitization. I: Background, procedures, and uncontrolled reports of individual treatment. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status*. New York: McGraw-Hill, 1969. (b)

Paul, G. L. Outcome of systematic desensitization. II: Controlled investigations of individual treatment, technique variations, and current status. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status*. New York: McGraw-Hill, 1969. (c)

Paul, G. L. Physiological effects of relaxation training and hypnotic suggestion. *Journal of Abnormal Psychology,* 1969, **74**, 425–437. (d)

Polin, A. T. The effect of flooding and physical suppression as extinction techniques on an anxiety-motivated avoidance locomotor response. *Journal of Psychology,* 1959, **47**, 253–255.

Prochaska, J. O. Symptom and dynamic cues in the implosive treatment of test anxiety. *Journal of Abnormal Psychology,* 1971, **77**, 133–142.

Rachman, S. Studies in desensitization, II: Flooding. *Behaviour Research and Therapy,* 1966, **4**, 1–6.

Rachman, S. Systematic desensitization. *Psychological Bulletin,* 1967, **67**, 93–103.

Rachman, S. Treatment by prolonged exposure to high intensity stimulation. *Behaviour Research and Therapy,* 1969, **7**, 295–302.

Schachter, S., & Singer, J. E. Cognitive, social, and physiological determinants of emotional state. *Psychological Review,* 1962, **69**, 379–399.

Solomon, R. L., Kamin, L. J., & Wynne, L. C. Traumatic avoidance learning: The outcomes of several extinction procedures with dogs. *Journal of Abnormal and Social Psychology,* 1953, **48**, 291–302.

Solomon, R. L., & Wynne, L. C. Traumatic avoidance learning: The principles of anxiety conservation and partial irreversibility. *Psychological Review,* 1954, **61**, 353–385.

Stampfl, T. G. Implosive therapy, Part I: The theory. In S. G. Armitage (Ed.), *Behavior modification techniques in the treatment of emotional disorders.* Battle Creek, Mich.: V. A. Publication, 1967.

Stampfl, T. G., & Levis, D. J. Implosive therapy. In R. M. Jurjevich (Ed.), *Handbook of direct and behavior psychotherapies.* Englewood Cliffs: Prentice-Hall, 1966.

Stampfl, T. G., & Levis, D. J. Essentials of implosive therapy: A learning-theory-based psychodynamic behavioral therapy. *Journal of Abnormal Psychology,* 1967, **72**, 496–503. (a)

Stampfl, T. G., & Levis, D. J. Phobic patients: Treatment with the learning theory approach of implosive therapy. *Voices: The Art and Science of Psychotherapy,* 1967, **3**, 23–27. (b)

Stampfl, T. G., & Levis, D. J. Implosive therapy—a behavioral therapy? *Behaviour Research and Therapy,* 1968, **6**, 31–36.

Stampfl, T. G., & Levis, D. J. Learning theory: An aid to dynamic therapeutic practice. In L. D. Eron & R. Callahan (Eds.), *The relation of theory to practice in psychotherapy.* Chicago: Aldine, 1969.

Staub, E. Duration of stimulus-exposure as determinant of the efficacy of flooding procedures in the elimination of fear. *Behaviour Research and Therapy,* 1968, **6**, 131–132.

Underwood, B. J. *Psychological research.* New York: Appleton-Century-Crofts, 1957.

Watson, J. P., & Marks, I. M. Relevant and irrelevant fear in flooding—A crossover study of phobic patients. *Behavior Therapy,* 1971, **2**, 275–293.

Wilkins, W. Desensitization: Social and cognitive factors underlying the effectiveness of Wolpe's procedure. *Psychological Bulletin,* 1971, **76**, 311–317.

Willis, R. W., & Edwards, J. A. A study of the comparative effectiveness of systematic desensitization and implosive therapy. *Behaviour Research and Therapy,* 1969, **7**, 387–395.

Wilson, G. D. Efficacy of "flooding" procedures in desensitization of fear: A theoretical note. *Behaviour Research and Therapy,* 1967, **5**, 138.

Wilson, G. T., & Davison, G. C. Process of fear reduction in systematic desensitization: Animal studies. *Psychological Bulletin,* 1971, **76**, 1–14.

Wolpe, J. *Psychotherapy by reciprocal inhibition.* Stanford: Stanford University Press, 1958.

Wolpe, J. *The practice of behavior therapy.* New York: Pergamon Press, 1969.

Wolpe, J., & Lazarus, A. A. *Behavior therapy techniques: A guide to the treatment of neuroses.* Oxford: Pergamon Press, 1966.

Wolpin, M., & Raines, J. Visual imagery, expected roles and extinction as possible factors in reducing fear and avoidance behavior. *Behaviour Research and Therapy,* 1966, **4**, 25–37.

(Received May 11, 1972)

Psychological Bulletin
1973, Vol. 79, No. 5, 335-340

CRITERION SHIFTS IN RECALL [1]

ALICE F. HEALY AND CINDY JONES [2]

Rockefeller University

Criterion shifts in recall are examined in light of clarifications of the distinction between Type I and Type II signal detection analyses and of the concept of criterion. With a new index of the criterion, Murdock's evidence for criterion shifts in recall is found rather to be evidence that subjects hold their criteria constant across different retention intervals. However, with the new criterion index, criterion shifts are found in intelligibility experiments and in the short-term recall experiment by Bjork and Healy. In these studies, subjects' criteria became more lax with increasing task difficulty. A possible strategy used by the subjects is proposed to explain this finding as well as the finding that the subjects shift their criteria in Bjork and Healy's task but not in Murdock's. Finally, this analysis is applied to recall studies in which the subject is allowed to leave blanks.

Recent applications of signal detection theory to the study of recall have been concerned with the problem of criterion shifts. The existence of such shifts has been the subject of much debate (Banks, 1970; Banks, 1971; Bernbach, 1971; Murdock, 1966), but this debate has been clouded by confusion on two basic points. First, not all authors have correctly distinguished between standard signal detection analysis (Type I) and Type II analysis. Second, most authors, even those of the original signal detection studies (e.g., Swets, Tanner, & Birdsall, 1961), have been inconsistent in their use of the term "criterion." This paper attempts to clarify the concept of criterion and reviews the evidence for criterion shifts in the light of this clarification.

TYPE I VERSUS TYPE II ANALYSES

The distinction between Type I and Type II signal detection analyses has been mistakenly equated (Banks, 1970, 1971) with the distinction between the binary decision and confidence rating procedures. Clarke, Birdsall, and Tanner (1959) defined Type I analysis as that applied when the subject evaluates external stimuli provided by the experimenter, while Type II analysis is applied when the subject evaluates his own responses. In either case, the subject's evaluation may be a binary decision or a confidence rating. In the Type I case of evaluating an external stimulus, the subject may use the binary decision procedure by responding "signal/noise" in a signal detection paradigm or "old/new" in a recognition task; alternatively, he may use the confidence rating procedure by giving his level of confidence that the signal was present or that a recognition test item was an old item. For the Type II situation, in which the subject evaluates a response he has just made on a test trial, he may give a binary decision by responding "correct/error" or he may use a confidence rating to indicate his judgment of the correctness of his response. Clearly, confidence ratings are simply extensions of binary decisions; with ratings the subject can give a multivalued evaluation of a stimulus or of his response, whereas with the binary decision procedure he is restricted to a two-valued judgment. Thus, binary decision yields a single point on the receiver-operating-characteristic (ROC) curve, while a confidence rating procedure with n rating levels yields $n - 1$ points.

The distinction between Type I and Type II signal detection analyses is subject to an even more subtle confusion, however, than that of equating it with the difference between the binary decision and confidence rating procedures. Clarke et al. (1959) referred to Type I analysis as stimulus conditional and to Type II analysis as response conditional. As mentioned above, this distinction can be based on whether the subject evaluates external stimuli or his own responses. However, the distinction can also be based on how the signal detection analysis is done: the ROC curves can be stimulus conditional or response conditional depending upon how the hit (or false alarm) rate is conditionalized. Specifically, the hit rate can be the probability of a certain evaluation by the subject given that the stimulus was a signal in a signal detection task or given that it was old in a recognition task. This is stimulus-conditional analysis. Alternatively, the hit rate can be the probability of a certain evaluation by the subject given that his current response was correct. This is response-conditional analysis. It is often assumed that these two bases for distinguishing Type I and Type II analyses, that is, what the subject does and how his data are analyzed, are necessarily equivalent. However, this assumption is open to debate.

---

[1] This research was supported in part by Public Health Service Grants GM01789 and GM16735 to the Rockefeller University. The authors are indebted to the members of the Rockefeller University Mathematical Psychology Laboratory for their contributions to this work.

[2] Requests for reprints should be sent to Alice F. Healy, the Rockefeller University, New York, New York 10021.

For instance, in the recognition experiment by Donaldson and Murdock (1968), the subject was asked to decide whether each stimulus was old or new and to give a confidence rating from 1 to 5 for his response. If Type I and Type II analyses are held to be defined by what the subject does, then this experiment must be treated with Type II analysis; but if Type I and Type II analyses are held to be distinguished by the method of analysis used, then this experiment can be handled by either analysis.

As a matter of fact, Donaldson and Murdock applied Type I analysis. They treated their data as a 10-point rating by the subject of his confidence that the stimulus was old; thus, an "old-5" response was taken as a rating of 10, while a "new-5" response was counted as a rating of 1. This enabled them to take their hit rate to be the probability that the subject emitted rating i ($i = 1, \ldots, 10$) given that the stimulus was old, rather than the probability that the subject emitted rating i ($i = 1, \ldots, 5$) given that his response was correct (i.e., was a hit or a correct rejection).

Since Type II analysis is response conditional, rather than stimulus conditional, it yields information not about the subject's response generation but only about his response evaluation. Thus, a Type II analysis of a recall task does not directly reveal the state of the subject's memory for the items but instead illuminates the evaluation process the subject applies to his recall responses. Values of d' and shifts in criteria revealed by Type II analysis reflect only this evaluation process, not the subject's memory state.

INDEX OF CRITERION

Inconsistent use of the term criterion has led to problems in determining whether criterion shifts exist in the recall situation. Part of the difficulty stems from the fact that instructions to the subject about what type of criterion to maintain vary over different experiments. The subject may be told explicitly to regulate his hit and false alarm rates (Egan, Schulman, & Greenberg, 1959), or to categorize his responses according to a posteriori probability estimates (Clarke, 1960), or to consider a payoff matrix when making his responses (Swets et al., 1961). These different instructions necessarily define different criterion indices. However, in most experiments, the subject is told no more than to be "strict" or "lax" in making his responses.

The most plausible psychological meaning of these instructions seems to involve the proportion of hits to false alarms that the subject allows himself to make. If instructed to maintain a strict criterion, the subject then allows himself very few false alarms relative to the number of hits he makes, whereas instructions to operate with a lax criterion indicate to the subject that he may allow many false alarms relative to the number of his hits. Most authors who employ a signal detection analysis identify the criterion as beta, the cutoff point value on the likelihood ratio scale. They characterize the criterion as "strict" or "lax" depending on whether beta has a high or low value, and for a constant probability of signal beta does indicate how many false alarms the subject allows himself to make relative to the number of his hits. However, when the probability of signal changes, beta no longer reflects the ratio of hits to false alarms. In fact, the cutoff point can remain constant while the subject's decisions become more strict or more lax in terms of the number of false alarms he allows for each hit that he makes; conversely, with a change in probability of signal, the cutoff point can change even though the subject is not being more strict or more lax. In view of this inconsistency, it would seem advantageous to find a new index of the criterion that reflects the ratio of hits to false alarms more directly.

If the subject is considered to be keeping his criterion constant as long as his ratio of hits to false alarms remains constant, then constant criterion means $P(H)/P(FA) = k$. For Type I analysis with a binary decision procedure, $P(H)$ is the joint probability of signal plus noise occurring and the subject saying "yes," so that $P(H) = P(SN \cdot Y)$. Similarly, $P(FA) = P(N \cdot Y)$ (Swets et al., 1961). Hence, constant criterion means

$$P(SN \cdot Y)/P(N \cdot Y) = k.$$

Since for any two events A and B,

$$P(A \cdot B) = P(B|A) \cdot P(A),$$

constant criterion also implies that

$$[P(Y|SN)/P(Y|N)] \cdot [P(SN)/P(N)] = k.$$

Therefore, if $P(SN)/P(N)$ decreases, then $P(Y|SN)/P(Y|N)$—the ratio of the hit rate to the false alarm rate, or HR/FAR—must increase, given a constant criterion. A change in HR/FAR implies that the likelihood ratio cutoff point, or beta, is also changed. Thus, if the probability of a signal changes during a task, the subject's cutoff point will also change if he is maintaining a constant criterion as defined above. This prediction that beta should change with changes in probability of signal is consistent with the prediction made by the standard calculation of the optimal value of beta (Swets et al., 1961): $\beta = V \cdot [P(N)/P(SN)]$, where V is the factor of costs and values. Therefore, those authors who contend that the subject should hold beta constant despite changes in $P(SN)$ are demanding that the subject behave in a nonoptimal manner. If, instead, the subject is expected to hold $P(H)/P(FA)$ constant, he can perform optimally.

The same argument holds for the case of n confidence ratings and Type II analysis (as well as for Type I confidence rating and Type II binary decision procedures). If the subject employs a constant criterion for his highest confidence rating, R_n, then $P(C \cdot R_n)/P(E \cdot R_n) = k$, where C and E represent corrects and errors, respectively. (For a lower rating, R_j,

$$P[C \cdot (R_n \vee R_{n-1} \vee \ldots \vee R_j)]/ P[E \cdot (R_n \vee R_{n-1} \vee \ldots \vee R_j)] = k_j.)$$

Thus, as above,

$$[P(R_n|C)/P(R_n|E)] \cdot [P(C)/P(E)] = k.$$

In a recall task, $P(C)/P(E)$ decreases with increasing retention interval; hence $P(R_n|C)/P(R_n|E)$, which is HR_n/FAR_n, increases with increasing retention interval, given a constant criterion. Again a change in HR_n/FAR_n—and hence a change in beta$_n$—does not constitute a criterion shift.

What measure, then, reflects criterion shifts? For Type I analysis with binary decision, a constant criterion means

$$P(SN \cdot Y)/P(N \cdot Y) = k.$$

Equivalently,

$$[P(SN|Y/\mathrm{P}(N|Y)] \cdot [P(Y)/P(Y)] = k.$$

Since $P(N|Y) = 1 - P(SN|Y)$, then

$$P(SN|Y) = k/(k + 1) = k'$$

for a given criterion. Thus, the a posteriori measure, $P(SN|Y)$, is an index of the criterion. As $P(SN|Y)$ changes or remains constant, so does the subject's criterion. Similarly, with Type II analysis and confidence ratings, $P(C \cdot R_n)/P(E \cdot R_n) = k$ for a constant criterion. Thus,

$$[P(C|R_n)/P(E|R_n] \cdot [P(R_n)/P(R_n)] = k,$$

and since $P(E|R_n) = 1 - P(C|R_n)$, then $P(C|R_n) = k/(k + 1) = k'$ for a given criterion. $P(C|R_n)$, therefore, indexes criterion changes. (Likewise, $P(C|R_n \vee R_{n-1} \vee \ldots \vee R_j)$ is the index for the jth criterion.)

EVIDENCE FOR CRITERION SHIFTS

With this criterion index in hand, what evidence can be found for criterion shifts in

TABLE 1

Criterion Indices for Pollack and Decker (1958)

Criterion index	Signal-to-noise ratio		
	−9 db.	−12 db.	−15 db.
$P(C\|R_6)$	.95	.86	.90
$P(C\|R_6 \vee R_5)$	.93	.81	.84
$P(C\|R_6 \vee \cdots \vee R_4)$	.89	.70	.74
$P(C\|R_6 \vee \cdots \vee R_3)$	.85	.67	.60
$P(C\|R_6 \vee \cdots \vee R_2)$	.88	.60	.49

recall? The classic study on this question is Murdock's (1966). On each trial the subjects were shown a list of paired associates and were tested for recall of one of the items on this list. The subjects also gave confidence ratings of their responses. Murdock employed Type II analysis and found an increase in beta with increasing retention interval (serial position) for the highest confidence rating, the only rating he analyzed. However, he did not find evidence for a criterion shift as defined above, for $P(C|R_n)$ was constant across retention intervals. Therefore, Murdock's finding of an increasing beta across retention intervals can be viewed as the natural consequence of the subjects' attempts to keep their criteria constant, rather than as an artifact of his experimental design as Banks (1970) argues. Since Murdock's results indicate that subjects can maintain a constant criterion over task difficulty, it is reasonable to ask if there is evidence that subjects ever shift their criteria over tasks of varying levels of difficulty.

Intelligibility experiments provide a paradigm analogous to recall tasks in which to study criterion shifts. Such experiments, like recall, are generally treated with Type II analysis since the number of stimulus alternatives involved is greater than two. The general paradigm requires the subject to identify which member of a set of test words was presented and to rate his confidence in his response. Pollack and Decker (1958) used such a paradigm and varied speech-to-noise ratio in order to manipulate task difficulty. They observed a constant $P(C|R_n)$ across different speech-to-noise ratios, which indicates that their subjects were able to hold their highest criteria constant. The indices for the remaining criteria can be calculated from the conditional proportions $P(C|R_i)$ and the proportions $P(R_i)$, both of which were presented graphically by Pollack and Decker, by means of the following equation:

$$P(C|R_{k+1} \vee R_k) = \frac{P(C|R_{k+1}) \cdot P(R_{k+1}) + P(C|R_k) \cdot P(R_k)}{P(R_{k+1}) + P(R_k)}.$$

The values of these indices are shown in Table 1. As the table indicates, these lower criterion indices did not stay constant across task difficulty; rather, they seem to decrease with decreasing signal-to-noise ratio or increasing task difficulty. Pollack and Decker, then, like Murdock, found no evidence for a shift of the highest criterion, but they did find shifts of the lower criteria.

In a subsequent experiment Decker and Pollack (1958) considered the same problem but presented words under several filtering conditions (low-pass frequency cutoffs) rather than at several speech-to-noise ratios. Decker and Pollack did not include the figures necessary to calculate the lower criterion indices, but they did find that $P(C|R_n)$ decreased with decreasing cutoff frequency, or increasing task difficulty, indicating a shift of the highest criterion toward increasing laxness with increasing task difficulty. Decker and Pollack reasoned that a shift of the highest criterion occurred in this experiment with filtered noise but not in the earlier experiment with varying speech-to-noise ratios because the qualitative features and the average intelligibility at the various speech-to-noise ratios were not markedly different, while the "sound" of the words at the various filtering levels did contrast noticeably.

A later study by Clarke (1960) furnishes evidence for criterion shifts across different signal-to-noise ratios. Clarke varied the number of stimulus alternatives as well as the signal-to-noise ratio. He did not provide the proportions $P(R_i)$ needed to compute the indices for the lower criteria, but he did find $P(C|R_n)$ to be directly related to speech-to-noise ratio and inversely related to message set size. Again, the highest criterion shifted toward increasing laxness with increasing task difficulty. Clarke explained this finding by hypothesizing that the subject rates a given item relative to the other items in the same test block rather than to all other items heard in a session. Indeed, items were blocked by message set size and signal-to-noise ratio in his experiment and similarly blocked by task difficulty in the two experiments previously discussed.

Although criterion shifts seem to be evident in intelligibility experiments, the question remains whether there is evidence for criterion shifts in recall experiments. Later papers (e.g., Bernbach & Bower, 1970) reporting replications of Murdock's finding of a constant d' and an increasing beta across retention intervals have not provided the data necessary to compute the criterion indices. However, such data were available to us from an experiment performed by Bjork and Healy at Rockefeller University.[3] This experiment tested short-term recall at various retention intervals. On each trial the subject was shown four letters followed by 3, 8, or 18 digits. As in Murdock's experiment, retention interval was randomized across trials; however, here retention interval was determined by the number of digits intervening between presentation and test rather than by serial position. Moreover, on each trial of this experiment, the subject was forced to recall all four letters in their proper order, while Murdock tested only one item per trial. As in Murdock's experiment, the subject also assigned to each item recalled a rating indicating how confident he was that it was correct. A 3-point scale was employed here; Murdock used a 6-point scale. Each of 12 subjects participated in two sessions of this experiment with 72 trials per session.

Table 2 shows the criterion indices as computed for the two criteria separating the three confidence ratings for each of the 12 subjects at each of the three retention intervals. A repeated-measures analysis of variance (Winer, 1971) indicated that the null hypothesis of a constant criterion across retention intervals could be rejected for each criterion ($F = 4.00$, $df = 2/22$, $p < .05$, for the higher criterion; $F = 21.67$, $df = 2/22$, $p < .01$, for the lower criterion). As in the previous experiments where criterion shifts were found, the subject's criteria became more lax with increasing task difficulty.

TABLE 2

CRITERION INDICES FOR BJORK AND HEALY (1970)

Subject	$P(C \mid R_3)$			$P(C \mid R_3 \vee R_2)$		
	Number of digits intervening			Number of digits intervening		
	3	8	18	3	8	18
1	.88	.77	.70	.76	.65	.51
2	.97	.84	.88	.96	.82	.87
3	.69	.70	1.00	.65	.53	.68
4	.98	.95	.67	.97	.88	.66
5	.92	.68	.51	.92	.65	.43
6	.96	.89	.87	.93	.83	.80
7	.88	.80	.75	.86	.70	.65
8	.91	.88	.63	.93	.80	.54
9	.80	.60	.88	.73	.40	.39
10	.95	.98	1.00	.94	.92	.93
11	.87	.74	.54	.86	.67	.43
12	.86	.64	.54	.84	.58	.44

DISCUSSION AND CONCLUSIONS

Why do we find evidence for criterion shifts in Bjork and Healy's data but not in Murdock's? If subjects shift their criteria as a result of a tendency to rate a given item relative to other items of the same retention interval rather than relative to all other items, they would be able to do so only if items from the different retention intervals were blocked or were highly discriminable. In both of these recall studies the various retention intervals were not blocked but were presented randomly throughout the session. However, it could be argued that retention interval was blocked to some extent in the Bjork and Healy experiment since on every trial the four letters tested, each of which was given a separate confidence rating by the subject, necessarily came from the same retention interval. In addition, trials at the different retention intervals were probably more discriminable in the Bjork and Healy experiment, where retention interval depended on number of intervening digits, than in Murdock's experiment, where retention interval depended on serial position.

---

[3] E. L. Bjork & A. F. Healy. Intra-item and extra-item sources of acoustic confusion in short-term memory. In, *Communications in mathematical psychology.* (Technical Report series) New York: Rockefeller University, 1970.

Such evidence for criterion shifts does not bear solely on experiments which employ confidence ratings. These findings have an important implication for experiments in which subjects are not forced to respond to every item but instead are free to leave blanks when uncertain.[4] In these experiments $P(C)$ for overt responses can be seen as analogous to $P(C|R_n)$ so that a change in $P(C)$ for overt responses would be evidence for a criterion shift. (See Lockhart, 1969, for a similar argument.) If such criterion shifts do occur and go undetected, the result could be a distorted picture of memory loss with increasing retention interval. $P(C)$ for all responses might be inflated at the longer intervals due to a more lax criterion there. Thus, although Type II analysis of recall cannot yield any new information about the subject's memory for the given items in a task, Type II analysis can illuminate the important evaluation processes which go on in recall and which may affect the measure of memory loss in such experiments.

---

[4] For example, W. K. Estes. On the source of acoustic confusions in short-term memory for letter strings. In, *Communications in mathematical psychology*. (Technical report series) New York: Rockefeller University, 1970.

REFERENCES

Banks, W. P. Signal detection theory and human memory. *Psychological Bulletin,* 1970, **74**, 81–99.

Banks, W. P. Confidence-rated recall, d_r, and tests of Bernbach's finite-state theory in recall. *Psychological Bulletin,* 1971, **76**, 151–152.

Bernbach, H. A. Invariance of d^* in memory: Response to Banks. *Psychological Bulletin,* 1971, **76**, 149–150.

Bernbach, H. A., & Bower, G. H. Confidence ratings in continuous paired-associate learning. *Psychonomic Science,* 1970, **21**, 252–253.

Clarke, F. R. Confidence ratings, second-choice responses and confusion matrices in intelligibility tests. *Journal of the Acoustical Society of America,* 1960, **32**, 35–46.

Clarke, F. R., Birdsall, T. G., & Tanner, W. P., Jr. Two types of ROC curves and definitions of parameters. *Journal of the Acoustical Society of America,* 1959, **31**, 629–630.

Decker, L., & Pollack, I. Confidence ratings and message reception for filtered speech. *Journal of the Acoustical Society of America,* 1958, **30**, 432–434.

Donaldson, W., & Murdock, B. B., Jr. Criterion change in continuous recognition memory. *Journal of Experimental Psychology,* 1968, **76**, 325–330.

Egan, J. P., Schulman, A. I., & Greenberg, G. Z. Operating characteristics determined by binary decisions and by ratings. *Journal of the Acoustical Society of America,* 1959, **31**, 768–773.

Lockhart, R. S. Retrieval asymmetry and the criterion problem in cued recall. *Journal of Experimental Psychology,* 1969, **81**, 192–194.

Murdock, B. B., Jr. The criterion problem in short-term memory. *Journal of Experimental Psychology,* 1966, **72**, 317–324.

Pollack, I., & Decker, L. Confidence ratings, message reception, and the receiver operating characteristic. *Journal of the Acoustical Society of America,* 1958, **30**, 286–292.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. Decision processes in perception. *Psychological Review,* 1961, **68**, 301–340.

Winer, B. J. *Statistical principles in experimental design.* (2nd ed.) New York: McGraw-Hill, 1971.

(Received May 15, 1972)

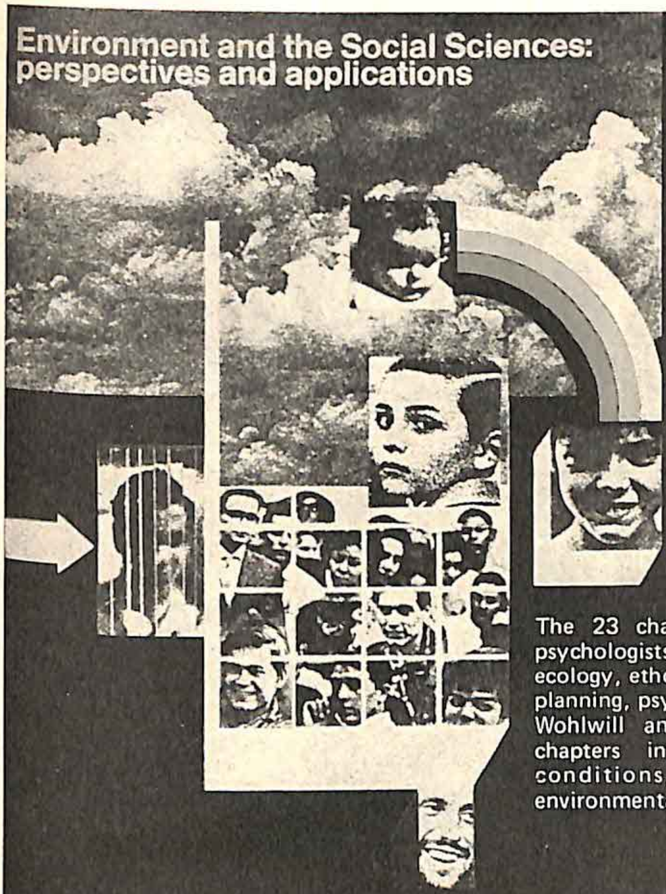
ERRATA

In the article "Defining and Testing Hypotheses in Multidimensional Contigency Tables" by Juliet Popper Shaffer in the February 1973 issue, the concluding parentheses of Footnote 3 on page 128 should read as follows:

(Bishop, 1969; Goodman, 1970, 1971a).

In addition, the first sentence under the heading *Log-Linear Model* on page 128 should read as follows:

The log-linear model, with properties investigated extensively by Birch (1963), Bishop (1969), Goodman (1964a, 1969b, 1970, 1971a), and Haberman (1970), provides. . . .



Environment and the Social Sciences: perspectives and applications

represents a dawning realization that psychology has a substantial contribution to make to a better understanding of our present-day environmental problems. . .

The 23 chapters include investigations not only by psychologists, but by professionals in architecture, ecology, etholcgy, geography, resource management and planning, psychiatry, and sociology. Editors Joachim F. Wohlwill and Daniel H. Carson have grouped the chapters into three parts: general environmental conditions, special environmental settings, and environmental decision making.

This new book is based on a series of symposia, round table discussions, and paper sessions on environmental topics presented at the APA Annual Convention in September 1970. Additional material was solicited from specialists working on significant problems not dealt with at that convention, while a number of previously published papers were also selected to strengthen coverage of certain topics. Printed on recycled paper.

300 pages All orders must be prepaid. $5.50

ORDER FORM

______ total enclosed for

______ copies of *Environment and the Social Sciences: perspectives and applications.*

Name______________________________

Address ____________________________

__________________________ Zip ________

Order Department
AMERICAN PSYCHOLOGICAL ASSOCIATION
1200 Seventeenth Street, N.W.
Washington, D.C. 20036

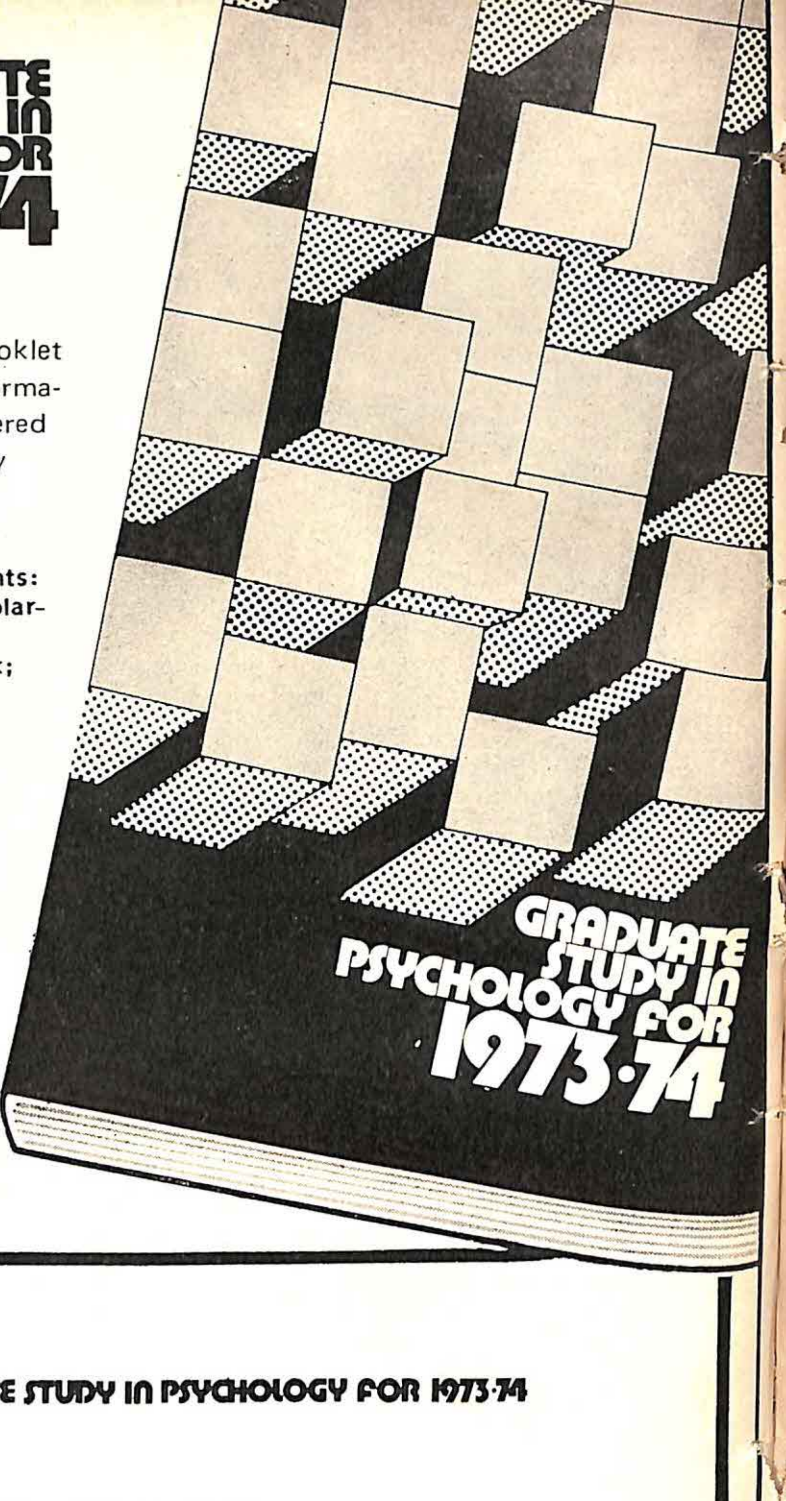
GRADUATE STUDY IN PSYCHOLOGY FOR 1973-74

The new sixth edition of this booklet provides relevant up-to-date information on 395 separately administered graduate programs in psychology in 315 universities and colleges.

Areas covered include enrollments: application for fellowships, scholarships, or assistantships; types of assistantships and hours of work; government stipends; and postdoctoral arrangements. Current information has been supplied by the department chairman, his or her representative, or the program director.

$2 per copy

All orders amounting to $15 or less must be prepaid.


GRADUATE STUDY IN PSYCHOLOGY FOR 1973-74

☐ total enclosed for

☐ copies of **GRADUATE STUDY IN PSYCHOLOGY FOR 1973-74**

Order Department
AMERICAN PSYCHOLOGICAL ASSOCIATION
1200 Seventeenth Street, N.W.
Washington, D. C. 20036

Name ____________________

Address ____________________

____________________ Zip ____________

VOL. 79, NO. 6 JUNE 1973

Psychological Bulletin

© 1973 by the American Psychological Association, Inc.

LONG-TERM PSYCHOLOGICAL EFFECTS OF LSD[1]

SPENCER A. McWILLIAMS[2] AND RENÉE J. TUTTLE

University of Arizona

The literature reporting lasting adverse psychological reactions to LSD is reviewed, along with other relevant reports where LSD users were followed. The danger of long-lasting psychological damage is low when the drug is used by emotionally stable individuals in secure, controlled settings, but persons with psychiatric disturbance, unstable personalities, and current crisis situations have experienced pathological behavior temporally related to drug ingestion. Adverse reactions were not reported in well-controlled studies with normal subjects, indicating the drug's relative safety for continued research.

Although lysergic acid diethylamide (LSD) was discovered to have hallucinogenic powers in 1943, drugs with similar makeup and pharmacological effects have been used for many years, especially by American Indians in their religious rituals (Pelner, 1967). Since the synthesis of LSD, these drugs have become a subject of empirical study among behavioral scientists studying the effects of such agents, their purported ability to induce schizophrenic states, and their use in psychotherapy. During the 1950s a number of research projects investigating the clinical effects of LSD were undertaken (See Cohen, 1967; Cole & Katz, 1964; Freedman, 1968, 1969; Ludwig & Levine, 1966; Osmond, 1957; Unger, 1963, for reviews).

Many laymen became aware of the reported ability of LSD to induce altered perceptual states and intense positive personal experiences (e.g., Huxley, 1954). This "psychedelic" orientation (Osmond, 1957) led to use of LSD in the general population, particularly among the young. Although long-term reactions to LSD had occasionally been reported from medical administrations (Cohen, 1960), the increasing nonmedical usage prompted warnings about the possibility of adverse psychological effects resulting from LSD usage.

A pamphlet distributed by the State of New York warned that LSD can precipitate an acute psychotic episode, requiring extensive treatment and from which some persons may never recover (Narcotic Addiction Control Commission, 1968). Grinker (1963, 1964) warned against the use of LSD unrelated to scientific research and criticized the results of LSD therapy suggesting that LSD could induce harm. Louria (1967) associated use with liberalization of morals and permissiveness and criticized LSD because its users withdraw from productive activities in society. He reported that many patients were hospitalized with psychoses induced by LSD. Largely because of such warnings, possession of LSD is now illegal, and research on its pharmacological effects and use in psychotherapy has decreased.

The issue is a controversial one, however, and not all authorities agree with the above conclusions. Hollister (1967) stated that LSD does not induce permanent psychosis and explained that the reported psychoses were due likely to the attraction of such drugs to people who are already emotionally disturbed. Freedman (1968) concurred and said that a causal relation between emotional disorder and LSD

---

[1] Gratitude is expressed to Helen and Vincent Nowlis, University of Rochester, for their helpful comments on an earlier version of this paper.

[2] Requests for reprints should be sent to Spencer A. McWilliams, Department of Psychology, University of Arizona, Tucson, Arizona 85721.

Diary No 685

usage was not justified simply by the presence of emotional turmoil in concurrence with the drug.

The unfavorable publicity surrounding LSD usage has hampered research efforts. Dahlberg, Mechanek, and Feldstein (1968) surveyed 29 investigators doing research on LSD-type drugs. The 19 who responded indicated that unfavorable publicity made it difficult for them to recruit appropriate subjects and caused pressure to discontinue research with the drug. Mogar (1968), in announcing his decision to withdraw entirely from research with psychedelics, discussed what he considered an oppressive social climate surrounding the drugs, which hindered the investigator. He felt that prevailing attitudes made creative and objective research difficult, and open and rational discussion of the drugs was hindered because of biases generated by sensationalism and controversy. Because of the illegality of the drugs, subsequent nonmedical usage has gone "underground," making it difficult to locate people who would admit to taking the drugs and producing sampling biases in subject selection (Blum, 1964).

Since much of the public furor over use of LSD was generated by claims that its use can engender permanent harm, it is important to investigate the evidence reporting ill effects to determine how closely its usage is related to long-term psychological damage. Reviews report adverse reactions to LSD (Greenblatt & Shader, 1970; Schwartz, 1968; Smart & Bateman, 1967), but attempts have not been made to differentiate between various conditions under which the drug was taken. In addition, these reviews failed to relate incidences in which follow-ups did not report ill effects.

The present study reviews the literature on long-term adverse psychological effects. Such a reaction is defined as pathological behavior (including suicide, psychosis, and severe anxiety) lasting for at least 48 hours after ingestion of the drug, the time by which the pharmacologic effects have usually subsided (Hoffer, 1965). These criteria for adverse reactions are difficult to follow in some cases since some reports did not distinguish between a dysphoric drug experience and a long-term effect. In addition, investigators have defined adverse reactions in different ways. In one survey (Ungerleider, Fisher, Goldsmith, Fuller, & Forgy, 1968), for example, a physician's biases toward LSD led him to state: "I would consider *all* reactions to LSD as 'adverse' regardless of the immediate subjective response [p. 355]." Reports of negative reactions which subsided with drug effects were not included. The bulk of the literature in this area consists of reports of adverse reactions, but in addition, studies which reported the use of LSD and included a follow-up of the users were included.

LSD has typically been administered under one of three basic conditions: (*a*) to psychiatric patients in research or psychotherapy programs; (*b*) in unsupervised, nonmedical settings, and (*c*) to normal subjects in research programs. Since the setting under which such drugs are taken influences the outcome of the experience, the present discussion is divided accordingly. Most of the literature consists of case studies and statistical surveys which could easily be criticized on methodological grounds. There are a few unusually clean and well-planned studies (Ditman, Tietz, Prince, Forgy, & Moss, 1967; McGlothlin, Cohen, & McGlothlin, 1967, 1968; Savage, Fadiman, Mogar, & Allen, 1966; Shagass & Bittle, 1967) but these are exceptional.

PSYCHIATRIC PATIENTS IN PSYCHOTHERAPY AND RESEARCH

Much of the original research with LSD was in drug therapy. Procedures and programs devised for LSD therapy have appeared (Abramson, 1967; Shlein, 1968, Section 3).

Cohen (1960) sent a questionnaire to 62 investigators using LSD or mescaline and received 44 responses with reference to nearly 5,000 individuals and more than 25,000 administrations. He reported four suicide attempts among patients (1.2 per 1,000 cases) and concluded that a direct connection between LSD and suicidal behavior could be discerned in very few instances. The one completed suicide, previously discussed by Savage (1957), was a chronic, regressed, schizophrenic girl with a 10-year hospitalization who had undergone LSD therapy. When inadvertently allowed out of the hospital on a weekend pass, she committed suicide by throwing herself under a train.

Hoffer (1965), discussing Cohen's study, felt that the rate of suicide was very low considering the hopelessness of most of the patients and concluded that LSD may have decreased the rate since a high suicide rate could be expected in such a group. Smart and Bateman (1967), however, criticized Cohen's findings on the grounds that the rate of complications for the therapists who failed to respond may have been higher and that there was little follow-up of some of the patients. They reported 14 attempted and 6 successful suicides following Cohen's (1960) study, but concluded that it was difficult to attribute suicidal behavior to LSD since the cases were suicide-prone prior to drug use.

Malleson (1971) sent questionnaires to all physicians in the United Kingdom who were known to have used LSD on human subjects. Of the 74 physicians, 73 replied reporting on 4,300 patients with approximately 49,000 treatments, and 170 experimental subjects with 450 administrations. The reported suicide rate was .7 per 1,000 cases with three completed suicides and nine serious attempts. Eleven additional suicide attempts were reported, but no direct relation between the suicidal behavior and LSD ingestion could be determined. Response to this survey was nearly 100% and included approximately the same number of subjects but a larger number of administrations than Cohen's (1960) survey, thus speaking to the earlier criticism (Smart & Bateman, 1967) that the rate of complication was reduced due to the number of nonresponders.

Chandler and Hartman (1960) reported a case of suicide in a woman after her first treatment with LSD. She had a long history of alcoholism and narcotic addiction and had made three previous suicide attempts. Prior to the LSD session, she confessed to her therapist that she was planning to commit suicide on a particular night, and after an argument with her husband she made good her threat. This was the only adverse reaction in a project involving LSD therapy with 110 patients.

In Cohen's (1960) survey, eight instances (1.8 per 1,000 cases) of lasting psychotic episodes following LSD were reported, seven of these in patients undergoing therapy. Five had long-term courses while two went into remission within a few weeks. Malleson's (1971) survey reported 35 cases of psychosis (9 per 1,000), 19 of which recovered completely in less than 3 months. There were 10 cases of long-term psychosis, some of which were seen as potential psychotics who would have succumbed eventually regardless of LSD use. Details relating to the other eight psychotic episodes were unknown.

Cohen and Ditman (1962) reported five instances of psychotic breaks. They found underlying hysterical or paranoid personalities in each and concluded that the reactions were due to unskillful therapeutic management, permitting an upsurge of unconscious material which the patient was unable to handle. In a subsequent article (Cohen & Ditman, 1963) they reported several cases of prolonged psychotic decompensation, depressive reaction, and paranoid reaction in LSD therapy patients. In most of these cases psychological disturbances antedated the treatment, but the patients had been functioning outside a hospital. The authors concluded that the incidence of complications following LSD was probably infrequent, although unknown, and usually occurred in emotionally labile, hysterical, or paranoid personalities.

Fink, Simion, Haque, and Itil (1966) administered LSD to 65 psychotic subjects with an average hospitalization of 5.1 years. There were 158 administrations, and three prolonged reactions were noted, consisting of disturbances of mood, affect, and thought. The symptoms appeared to be an exaggeration of the patient's pretreatment state. The risk of negative reaction to LSD therapy was greater in subjects with emotional lability and psychotic features than in classical forms of schizophrenia.

In a follow-up of patients who underwent a single experience with LSD or mescaline, Savage, Savage, Fadiman, and Harman (1964) reported three negative incidents in questionnaire responses. One patient felt he did not receive enough support during the session and suffered from obsessive ruminations for 9 months following the session, after which he improved. Another patient who had a long pre-LSD history of depression, hospitalization, and suicide attempts made another unsuccessful suicide attempt 2 months after the LSD session. A third patient did not keep his follow-up appointments and entered another facility where a second LSD treatment was proposed. He had a psychotic break on the evening of the proposed second session, but recovered. However, 83% of the respondents claimed lasting benefits, with an increase from 76% at 3 months to 85% after 6 months. The authors made clinical ratings on 74 of these patients 6 months after the session and found that 60 improved, 13 had not changed, while 1 was changed for the worse, although after 6 months of therapy he returned to his original base line. These ratings were based on clinical interviews and psychological data including the MMPI (pre- and post-) and the Leary Interpersonal Check List.

All of the preceding studies can be criticized on methodological grounds. A major flaw in each is the absence of a non-LSD control group comparable to LSD-taking patients in pretreatment pathology. One attempt to introduce experimental controls was made by McGlothlin and Arnold (1971) in a 10-year follow-up of 247 patients of three physicians who had experimented with LSD between 1955 and 1961. The experimental sample was matched with a control group of non-LSD patients from the same therapists based on psychological tests and attitude measures. No significant differences were found between the control group and subjects who had taken LSD only in therapist-initiated sessions. Of the experimental subjects, 19 were hospitalized with psychosis after their LSD experiences, but only one case was reported as being related to LSD ingestion. This was a patient who received three LSD treatments, was hospitalized 1 week after the final drug experience, and was unable to work for 3 months.

The above reports, although methodologically weak, indicate that the danger of psychiatric disturbance resulting from LSD is slight when the drug is used with proper subjects in well-supervised settings. Thus, the evidence does not provide grounds for discontinuing LSD treatment. Additionally, other projects did not report serious, prolonged, adverse reactions. Amarel and Cheek (1965) studied verbal behavior under LSD by administering 100 micrograms and 200 micrograms of LSD 1 week apart to 10 male alcoholic patients. They reported a 1-week follow-up after the second session with no adverse reactions. Bhattacharya (1966) reported on 581 LSD treatment cases and 2,742 administrations. He reported no cases of either temporary or permanent psychosis, no suicide attempts, and no behavior so bizarre that it was beyond control of the therapist.

Savage et al. (1966) reported on 77 patients who had psychedelic therapy over a 10-month period. Two thirds of the patients resembled typical psychiatric outpatient cases, while one third were mildly neurotic. The session consisted of a 200–300 microgram dose of LSD and lasted an entire day. The therapist was present to provide companionship but not interpretation. Patients were followed up to 6 months and were assessed by psychological tests, clinical evaluations, staff ratings, and behavior change. Results were generally positive and no adverse reactions were reported even though some patients were followed for up to 1 year.

Shagass and Bittle (1967) studied the effects of a large single dose of LSD 6 months and 1 year following the session. Patients were matched with non-LSD controls on sex, education, marital status, diagnosis, and symptoms. In addition, the experimental patients were divided into two groups: responders ($n = 8$), who displayed insight under LSD; and nonresponders ($n = 12$). Patients were rated on amount of improvement through an interview, and the experimental group improved significantly more than the control group ($p < .05$ at 6 months, $p < .02$ at 1 year). More controls than LSD subjects required follow-up therapy. No prolonged adverse reactions were reported.

Although it is apparent that LSD can produce an exacerbation of symptoms or psychotic behavior in some disturbed psychiatric patients, the literature is far from conclusive on the role of the drug experience in pathological behavior. One of the major problems is the fact that base-rate statistics for suicidal behavior and psychosis among patients are not available. Sandison (1966) stated that both suicide and psychosis are possible complications of psychotherapy even when no drugs are used. Levine and Ludwig (1964) also felt the results support a conclusion that the drugs are safe and indicated that perhaps the incidence of negative reactions is just as high or higher for comparable patients in other active therapies. The result of several other reports (e.g., Buckman, 1968; Savage, 1968) in which LSD therapy produced positive results lends support to the contention that LSD can be a safe therapeutic instrument when properly administered to suitable subjects.

REACTIONS TO UNSUPERVISED DRUG USAGE

The majority of reported adverse reactions have been from instances where the drug was taken in unsupervised, nonmedical settings. Undoubtedly, most LSD ingestions have occurred under such conditions, but the exact number of administrations, purity, and dosage level remains unknown. The literature consists mostly of case reports and surveys rather than experimental studies.

Blumenfield and Glickman (1967) reported on 25 patients with a history of LSD use seen in the psychiatric emergency room of a hospital over a 10-month period. Thirteen were admitted with symptoms related to LSD and schizophrenic or "borderline schizophrenic" diagnoses. In addition, four suicide and two homicide attempts were noted. Seventy-two percent of the patients had psychiatric treatment prior to taking LSD, and 40% of these had psychiatric hospitalizations. The authors felt that the current hospitalizations were related to previous psychiatric difficulties and life situations rather than the effects of the drug. This conclusion was supported in a follow-up of 15 of these patients (Glickman & Blumenfield, 1967). It was found that the patients began LSD ingestion as an attempt to resolve a life crisis, and the stresses were the same types which often lead to hospitalization without drug use. The authors concluded not only that the psychoses and suicides would have occurred if the patients had not taken LSD, but the fantasy of "cure" through LSD may have helped to postpone the actual break.

A survey of negative side effects of LSD was reported by Ungerleider and his colleagues (Ungerleider, Fisher, & Fuller, 1966; Ungerleider, Fisher, Fuller, & Caldwell, 1968). Seventy patients whose diagnoses were related to LSD, seen over a 7-month period, were reviewed. Twenty-five of these were admitted to the hospital. Twenty-seven percent were diagnosed psychotic, 21% neurotic, and 18% character disorder. Ten percent had previous psychiatric hospitalizations and 27% had received psychiatric outpatient care. The 25 inpatients were compared with a "control" group of 25 frequent LSD users, a group of rehabilitated criminals, who were not hospitalized to determine correlates of adverse reactions. The groups were compared on demographic background, marital status, employment, police record, education, previous psychiatric care, and the MMPI. Significant differences were found for marital status (0% inpatients married versus 60% controls, $p < .001$), employment (20% inpatients working versus 70% controls, $p < .01$), and police records (8% inpatients versus 64% controls, $p < .001$). This latter finding was due to the fact that the control group was composed of rehabilitated criminals. The inpatient group displayed severe psychopathy on the MMPI, but similar findings were not present in the controls.

Ditman et al. (1967) were also interested in why some people had adverse reactions while others did not and may even have found LSD therapeutic. Subjects were 116 people who had taken LSD. Group 1 ($n = 52$) consisted of people who did not need psychiatric care. Most were students or working people. Group 2 ($n = 27$) consisted of psychiatric outpatients who applied for treatment as a result of LSD ingestion, and Group 3 ($n = 37$) was composed of psychiatric patients hospitalized because of LSD use. Groups were compared on the basis of an interview, demographic variables, drug history, and the DWM Card Sort Test, an instrument developed to assess LSD experiences. The employment status of subjects was assessed prior to hospitalization but after they had taken LSD. The percentage of subjects who were employed or were full-time students was significantly higher in Group 1 than in Group 3 ($p < .001$). Since this was not assessed until after the subjects had taken LSD, however, a cause–effect relation cannot be determined. On the DWM Card Sort Groups 2 and 3 reported more unpleasant feelings under LSD than Group 1 ($p < .01$) and more intense feelings of despair and hopelessness ($p < .05$). Items reflecting a a positive, beneficial experience were reported more frequently by Group 1. The authors concluded that unstable individuals who take LSD without support and protection are likely to find the experience disruptive, while the results in the nontreatment group supports the notion that LSD has a potential value for research in psychotherapy and esthetic appreciation.

Frosch, Robbins, and Stern (1965) studied 12 of 27 patients admitted to a hospital as a result of LSD ingestion over a 4-month period. Seven were diagnosed as panic reactions, three had a reappearance of the drug symptoms without reingestion of the drug, and three were psychotic. The panic reactions rarely led to hospitalization. Of the three with reappearing symptoms, two were diagnosed as chronic schizophrenics and the other was a borderline psychotic. The three with extended psychoses following LSD ingestion had clear long-standing schizophrenia: two were catatonic and one was paranoid. The schizophrenic patients developed a unique sense of "insight" under LSD and were subsequently unable to adapt to the environment. A second sample ($n = 22$) was taken from the same hospital and compared with the original group (Robbins, Frosch, & Stern, 1967). This group had more drug experiences but the symptoms still fell into the same three categories: (*a*) 11 panic reactions, (*b*) 8 symptom reappearances, and (*c*) 8 extended psychoses. The recurring symptoms were seen in repeating users when exposed to an anxiety-arousing situation and were felt to involve psychological as well as physiological factors. Contrary to earlier findings, three of the psychotics were seen as not psychotic before taking LSD, although suggestions of severe personality disturbance were present. The authors warned that adverse effects could not always be predicted from previous history.

Louria (1967) reported over 130 patients hospitalized with LSD-induced acute psychoses, or chronic schizophrenia, and schizoid personality exacerbated by LSD. Of these, 34.2% were reported as having underlying personality disturbance, and the author concluded that an increasing number of susceptible, unstable individuals were being influenced to use LSD.

Not all studies have indicated preexisting psychiatric difficulties in those becoming psychotic. Tietz (1967) reported a study of 49 patients admitted to an inpatient service with LSD-related complications. Eighty percent were from the lower socioeconomic class, and dosages from 100 to 2,000 micrograms were reported. Patients received psychological testing and had Bender-Gestalt tests within normal limits, but MMPI profiles similar to schizophrenics. Fifteen had acute panic reactions, 5 requiring hospital care, while 6 had recurring symptoms and 28 had extended psychoses. Conclusions are limited by the lack of information on premorbid adjustment, but none had previous psychiatric hospitalizations.

Many reports on LSD reactions have been case studies, which, although valuable for gathering data, fail to provide a basis for determining a relation between LSD and subsequent reactions since they lack experimental controls. Appropriate caution must be used in making conclusions from these studies.

Bowers, Chipman, Schwartz, and Dann (1967) reported three cases of adverse reactions to LSD, two requiring treatment. A 19-year-old college sophomore was admitted with depression and disorganization after having taken LSD. The history revealed a poor childhood, a previous long-term hospitalization, and an inability to relate to others. A 20-year-old medical student was seen as an outpatient for five sessions because of anxiety following LSD but was not considered seriously disturbed. The authors concluded that each patient had intense needs for interpersonal closeness and no access to meaningful emotional experiences. Muller (1971) reported three cases of prolonged LSD effects leading to hospitalization. All were released after 2 weeks, although two returned after a short period with a relapse of symptoms. Electroconvulsive treatments produced remission in these cases.

Two cases of psychosis due to LSD ingestion of sufficient severity to warrant hospital admission were reported by Hatrick and Dewhurst (1970). Only one dosage of LSD was involved in each case, and in one case it was administered to the patient surreptitiously. Both patients had well-adjusted premorbid personalities, no previous history of regular drug use, and both appeared normal for between 2 weeks and 2 months before the onset of psychotic effects. Although the symptoms abated within a number of weeks, the authors concluded that these cases were of sufficient severity to warrant the removal of LSD from use in psychotherapy. In a rebuttal, Malleson (1970) argued that the nature of the relation between the drug and illness was indeterminable, the pathology was more likely brought on by emotional upset in the patients, and the conclusion that such therapy should be discontinued was not justified on the basis of anecdotal information.

Efforts to determine the relation between LSD and psychiatric disturbance have also been weakened by unreliable and subjective psychiatric diagnoses. Often, reports appear to be based on value judgments, and patient attitudes and life-styles rather than standard diagnostic criteria. Rossi (1971) discussed the "psychedelic syndrome" as one form of adverse reaction characterized by preoccupation with mystical subjects, belief in astrology and extrasensory perception, and a rejection of societal values. He related LSD usage in such cases as an attempt to avoid the decision between social commitment and passivity. Glass and Bowers (1970) reported four LSD-related cases hospitalized by their families because of bizarre behavior, appearance, and withdrawn manner. The patients had mystical attitudes, belief in oriental religions, and perceived themselves as passive agents manipulated by cosmic forces. All were well oriented with a good memory for recent events, although recall of past events was vague, and visual hallucinations and paranoid delusions were present. In each case, heavy drug use and subsequent reactions appeared when the patients were facing heavy stress. The authors inferred that heavy drug use may have affected concurrently developing attitudes and values, thus facilitating the acquisition of mystical beliefs. A major difficulty with this interpretation is the fact that mystical beliefs are shared by a large number of young persons with no history of LSD use. It is equally possible that the adverse reactions were due simply to an inability to deal with stress situations. The diagnosis of unpopular attitudes and mystical beliefs as psychiatric symptoms is based on cultural value judgments rather than objective psychiatric criteria and only obfuscates understanding of drug effects.

Although it is evident that persons may develop serious psychological disturbances after taking LSD, a causal relation between the two events is most difficult to determine. In nearly every case where a prolonged reaction was noted, the history showed signs of psychological disturbances which antedated drug use. In addition, evidence indicates that many persons who take the drugs do so with a belief that it will help them solve their emotional problems. It is impossible to determine the number of unsupervised LSD ingestions, but even with the most conservative estimates it is obvious that many more persons are using the drug than are having negative effects. The most reasonable conclusion from the preceding section is that LSD can cause psychological difficulties in disorganized and disturbed individuals who ingest the drug in less than secure surroundings, without psychological support, and at a time when emotional problems or crises are at a peak.

Use with Experimental Subjects

Although many experiments studying the effects of LSD on normal subjects have been reported, most investigated the effects of acute intoxication on psychomotor or psychological tasks. Few of these studies followed the subject for any period of time after the drug experience. As a result, although many reports appear, few are relevant to the present review. Studies investigating LSD and including a follow-up are discussed first, followed by studies designed to explore long-term effects.

Abramson and his colleagues performed a long series of experiments assessing the effects of LSD on many psychological and physiological responses (e.g., Abramson, Jarvik, & Hirsch, 1955; Abramson, Jarvik, Kaufman, Kornetsky, Levine, & Wagner, 1955). They used many of the same subjects in over 150 experiments over a 10-year period. The records of six subjects were shown (Abramson & Rolo, 1965) and no adverse reactions were reported. The possibility of adverse reactions was discussed and it was felt that no danger existed as long as drug usage was supervised by a physician.

In an experiment comparing the LSD experience to schizophrenia, Langs and Barr (1968) gave LSD to 30 normal subjects. Eight subjects with anxious, manipulative, and hostile traits on predrug testing showed behavior similar to paranoid schizophrenics while under LSD, but no long-term adverse effects were found. In two studies involving 69 penitentiary inmates, Katz, Waskow, and Olsson (1968) gave 50 micrograms of LSD to 23 subjects and found either euphoric, dysphoric, or ambivalent responses. No adverse reactions were reported. Keeler (1965) gave psilocybin to 12 normal subjects, personnel at the research center where he worked, and found significantly higher scores on the schizophrenia and hypochondria scales of the MMPI from the pretest to a test taken while under the drug. Since the subjects had employment contact with the experimenter after the session, a long follow-up is assumed, but no adverse reactions were reported.

Welpton (1968) performed an intensive week-long study of LSD users and gave them LSD under his supervision. Extensive screening, testing, and interviewing were undertaken and poor family histories as well as some psychological maladjustment were found. "The most impressive finding was the relative infrequency of adverse reactions to LSD and other hallucinogens considering the degree of personality disturbances [p. 378]." In Cohen's (1960) original survey, one psychotic reaction in an experimental subject was reported. This was a man who was an identical twin of a hospitalized schizophrenic, and the reaction subsided within a few days. Pollard and Uhr (1965) reported on 5 years of experimentation using LSD and psilocybin with 80 college student subjects and reported no persistent ill effects, due to careful subject screening.

The low incidence of negative experiences among experimental subjects is even more impressive considering a series of findings reported by Cohen (1960). He related several studies in which the following reactions were attributed to the drug experience: migraine headaches, influenza, paraplegia, a suicide attempt, and a skin rash. All of these reactions occurred in control subjects who received water placebos. This finding indicates that hysterical, suggestible subjects may explain subsequent difficulties on the "drug" experience, indicating that reactions reported due to the drug may have been related instead to personality factors.

One of the most extensive and well-controlled programs to investigate the long-term effects of LSD was conducted by McGlothlin et al. (1967, 1968), assessing change due to LSD in personality, attitudes, values, interests, and performance among normals. Subjects ($n = 72$) were male graduate students who volunteered for the experiment without knowing it involved LSD. Their knowledge of LSD was determined and those with LSD experience, currently in psychotherapy, or psychoses in the family were screened out. Others were dropped because of interview impressions or doubtful MMPI profiles. Subjects were given a battery of tests assessing anxiety, personality, attitudes, esthetic sensitivity, and creativity. Then, they were divided into three groups of 24, matched with respect to knowledge of the drugs and expectations, and received the following drugs in a double-blind design: Experimentals, 200 micrograms of LSD; Control 1, 20 milligrams of amphetamine; and Control 2, 25 micrograms of LSD (below threshold dosage). Subjects met for three sessions and received the same drug and dosage each time. On 2-week and 6-month follow-ups, subjects were retested with the same instruments. Thirty-three percent of the experimental subjects had lower anxiety scores compared to 13% in Control 2 and 9% in Control 1 after 2 weeks. Experimental subjects had a significantly lower galvanic skin response to stress at the 6-month follow-up. Sixty-two percent of the experimental subjects reported an increased appreciation of music, which was supported by behavioral changes such as buying records and attending concerts. Other tests showed differences in favor of the experimental group but most failed to reach statistical significance. No adverse reactions were reported, although one subject was terminated due to an "unrealistic reaction" to the first drug dose.

Ditman, Hayman, and Whittlesey (1962) studied 74 subjects with LSD experiences from 6 months to 3½ years prior to the follow-up. Half the subjects were "normals" and half were psychiatric patients, and each received 100 micrograms of LSD. Questionnaires regarding the experience were completed at the follow-up and 66% reported some degree of benefit. Patients claimed more improvement than normals, although normals reported more pleasant LSD experiences. The authors concluded that if subjects' claims are valid, the LSD experience may be therapeutic in itself. No adverse reactions were reported.

An investigation of the effects of LSD on problem solving and creative abilities was reported by Harman, McKim, Mogar, Fadiman, and Stolaroff (1966). Subjects were 27 male professionals whose work required creative problem-solving abilities (engineers, physicists, mathematicians, architects, artists). They were found to be psychiatrically normal and motivated to discover and apply problem solutions within their professional capacity. They were asked to select a problem and encouraged to work on the problem under the drug. Subjects were given 200 micrograms of mescaline sulfate (equal to 100 micrograms of LSD) in groups of four, accompanied by two observers. They relaxed and listened to music for 3 hours, spent 1 hour on psychological testing, and worked on their problems for 3–4 hours. The psychometric data reported significant increases in creativity while under the drug and at the 2-week follow-up subjects reported significant changes in their functioning which related to their enhanced abilities during the drug experience. Subjects were seen 6 weeks after the session and no adverse reactions were reported.

The above studies indicate quite clearly that when psychologically well-adjusted individuals took LSD under controlled, secure, environmental situations where support was available, no long-term adverse reactions occurred.

CONCLUSIONS

The major conclusion of this review is that research on LSD, particularly its long-term psychological effects, is very poor. Most studies are based on subjective case reports, surveys, or used very weak experimental controls. The lack of comparable control groups and base-rate statistics for nondrug therapies makes interpretation of studies with patients very difficult. Any direct relation between drug use and subsequent behavior remains strictly conjectural. In many case studies, interpretations of behavior were often based on the biased judgment of the investigator rather than any standard criterion. Very few controlled studies investigating the areas of psychological functioning affected by LSD, especially those areas espoused by proponents of consciousness-expanding experiences, have appeared. The work of McGlothlin et al. (1967, 1968) represents the type of research with the greatest potential for generating scientific knowledge of the drug's effects.

Those adverse reactions which have occurred were most frequent among individuals who were emotionally disturbed, in crisis situations or insecure environments, and who took the drug in unsupervised settings. Psychotic episodes and suicidal behavior, while rare, have occurred in such circumstances. The few well-controlled studies indicate, however, that when the drug was administered to psychologically normal subjects under secure circumstances, lasting adverse reactions did not occur. Thus, the concerns which led to discontinuation of research are unsubstantiated, and the type of controlled research discussed above could be resumed without undue danger to subjects.

REFERENCES

ABRAMSON, H. A. (Ed.) *The use of LSD in psychotherapy and alcoholism.* New York: Bobbs-Merril, 1967.

Abramson, H. A., Jarvik, M. E., & Hirsch, M. W. Lysergic acid diethylamide (LSD-25): X. Effect on reaction time to auditory and visual stimuli. *Journal of Psychology,* 1955, **40**, 39–52.

Abramson, H. A., Jarvik, M. E., Kaufman, M. R., Kornetsky, C., Levine, A., & Wagner, M. Lysergic acid diethylamide (LSD-25): I. Physiological and perceptual responses. *Journal of Psychology,* 1955, **39**, 30–60.

Abramson, H. A., & Rolo, A. Lysergic acid diethylamide (LSD-25): XXXVII. Comparison with action of methysergide and psilocybin on test subjects. *Journal of Asthma Research,* 1965, **3**, 81–96.

Amarel, M., & Cheek, F. E. Some effects of LSD-25 on verbal communication. *Journal of Abnormal Psychology,* 1965, **70**, 453–456.

Bhattacharya, B. Lysergic acid diethylamide. *British Medical Journal,* 1966, **2**, 49.

Blum, R. The research enterprise and its problems. In R. Blum & associates (Eds.), *Utopiates: The use and users of LSD-25.* New York: Atherton Press, 1964.

Blumenfield, M., & Glickman, L. Ten months experience with LSD users admitted to a county psychiatric receiving hospital. *New York State Journal of Medicine,* 1967, **67**, 1849–1853.

Bowers, M., Chipman, A., Schwartz, A., & Dann, O. T. Dynamics of psychedelic drug abuse. *Archives of General Psychiatry,* 1967, **16**, 560–566.

Buckman, J. LSD in the psychotherapy of psychosomatic disorders the nature of unconscious material produced. In J. M. Shlien (Ed.), *Research in psychotherapy.* Vol. 3. Washington, D. C.: American Psychological Association, 1968.

Chandler, A. L., & Hartman, M. A. Lysergic acid diethylamide (LSD) as a facilitating agent in psychotherapy. *Archives of General Psychiatry,* 1960, **2**, 286–299.

Cohen, S. Lysergic acid diethylamide: Side effects and complications. *Journal of Nervous and Mental Disease,* 1960, **130**, 30–40.

Cohen, S. Psychotomimetic agents. *Annual Review of Pharmacology,* 1967, **7**, 301–318.

Cohen, S., & Ditman, K. S. Complications associated with lysergic acid diethylamide. *Journal of the American Medical Association,* 1962, **181**, 161–162.

Cohen, S., & Ditman, K. S. Prolonged adverse reactions to lysergic acid diethylamide. *Archives of General Psychiatry,* 1963, **8**, 475–480.

Cole, J. O., & Katz, M. M. The psychotomimetic drugs: An overview. *Journal of the American Medical Association,* 1964, **187**, 758–761.

Dahlberg, C. C., Mechanek, R., & Feldstein, S. LSD research: The impact of lay publicity. *American Journal of Psychiatry,* 1968, **125**, 685–689.

Ditman, K. S., Hayman, M., & Whittlesey, R. B. Nature and frequency of claims following LSD. *Journal of Nervous and Mental Disease,* 1962, **134**, 346–352.

Ditman, K. S., Tietz, W., Prince, B. S., Forgy, E., & Moss, T. Harmful aspects of the LSD experience. *Journal of Nervous and Mental Disease,* 1967, **145**, 464–474.

Fink, M., Simeon, J., Haque, W., & Itil, T. Prolonged adverse reactions to LSD in psychotics. *Archives of General Psychiatry,* 1966, **15**, 209–213.

Freedman, D. X. On the use and abuse of LSD. *Archives of General Psychiatry,* 1968, **18**, 330–347.

Freedman, D. X. The psychopharmacology of hallucinogenic agents. *Annual Review of Medicine,* 1969, **20**, 409–418.

Frosch, W. A., Robbins, C. S., & Stern, M. Untoward reactions to lysergic acid diethylamide (LSD) resulting in hospitalization. *New England Journal of Medicine,* 1965, **273**, 1235–1239.

Glass, G. S., & Bowers, M. B. Chronic psychosis associated with long-term psychomimetic drug use. *Archives of General Psychiatry,* 1970, **23**, 97–103.

Glickman, L., & Blumenfield, M. Psychological determinants of "LSD reactions." *Journal of Nervous and Mental Disease,* 1967, **145**, 79–83.

Greenblatt, D. J., & Shader, R. J. Adverse effects of LSD: A current perspective. *Connecticut Medicine,* 1970, **34**, 895–902.

Grinker, R. R. LSD: Editorial. *Archives of General Psychiatry,* 1963, **8**, 425.

Grinker, R. R. Bootlegged ecstasy. *Journal of the American Medical Association,* 1964, **187**, 768.

Harman, W. W., McKim, R. H., Mogar, R. E., Fadiman, J., & Stolaroff, M. J. Psychedelic agents in creative problem solving: A pilot study. *Psychological Reports,* 1966, **19**, 211–227.

Hatrick, J. A., & Dewhurst, K. Delayed psychosis due to LSD. *Lancet,* 1970, **2**, 742–744.

Hoffer, A. Lysergic acid diethylamide (LSD): A review of its present status. *Clinical Pharmacology and Therapy,* 1965, **6**, 183–255.

Hollister, L. E. Predictable and unpredictable drug responses in man: The nervous system. In G. Walstenhome & R. Porter (Eds.), *Drug responses in man.* Boston: Little, Brown, 1967.

Huxley, A. *The doors of perception.* New York: Harper, 1954.

Katz, M. M., Waskow, I. E., & Olsson, J. Characterizing the psychological state produced by LSD. *Journal of Abnormal Psychology,* 1968, **73**, 1–14.

Keeler, M. H. Similarity of schizophrenia and the psilocybin syndrome as determined by objective methods. *International Journal of Neuropsychiatry,* 1965, **1**, 630–634.

Langs, R. J., & Barr, H. L. Lysergic acid diethylamide (LSD-25) and schizophrenic reactions: A comparative study. *Journal of Nervous and Mental Disease,* 1968, **147**, 163–172.

Levine, J., & Ludwig, A. M. The LSD controversy. *Comprehensive Psychiatry,* 1964, **5**, 314–321.

Louria, D. B. LSD—perhaps a boon, nonmedically a danger. *Medical Tribune,* 1967, **8**, 18–19.

Ludwig, A. M., & Levine, J. The clinical effects of psychedelic agents. *Clinical Medicine,* 1966, **73**, 21–24.

Malleson, N. Reply to Hatrick & Dewhurst, delayed psychosis due to LSD. *Lancet,* 1970, **2**, 831.

Malleson, N. Acute adverse reactions to LSD in clinical and experimental use in the United Kingdom. *British Journal of Psychiatry,* 1971, **118,** 229–230.

McGlothlin, W. H., & Arnold, D. O. LSD revisited: A ten-year follow-up of medical LSD use. *Archives of General Psychiatry,* 1971, **24,** 35–49.

McGlothlin, W. H., Cohen, S., & McGlothlin, M. S. Long lasting effects of LSD on normals. *Archives of General Psychiatry,* 1967, **17,** 521–532.

McGlothlin, W. H., Cohen, S., & McGlothlin, M. S. Personality and attitude change in volunteer subjects following repeated administration of LSD. In J. M. Shlien (Ed.), *Research in psychotherapy.* Vol. 3. Washington, D. C.: American Psychological Association, 1968.

Mogar, R. E. Research in psychedelic drug therapy: A critical analysis. In J. M. Shlien (Ed.), *Research in psychotherapy.* Vol. 3. Washington, D. C.: American Psychological Association, 1968.

Muller, D. J. ECT in LSD psychosis: A report of three cases. *American Journal of Psychiatry,* 1971, **128,** 351–352.

Narcotic Addiction Control Commission. *Questions and answers: Barbiturates, amphetamines, LSD, marijuana, narcotics.* Albany: State of New York, 1968.

Osmond, J. A review of the clinical effects of psychotomimetic agents. *Annals of the New York Academy of Sciences,* 1957, **66,** 418–434.

Pelner, L. Peyote cult, mescaline hallucinations, and model psychosis. *New York State Journal of Medicine,* 1967, **67,** 2838–2843.

Pollard, J. C., & Uhr, L. *Drugs and phantasy.* Boston: Little, Brown, 1965.

Robbins, E., Frosch, W., & Stern, M. Further observations on untoward reactions to LSD. *American Journal of Psychiatry,* 1967, **124,** 393–395.

Rossi, G. V. LSD-A pharmacologic profile. *American Journal of Pharmacy,* 1971, **143,** 38–44.

Sandison, R. A. Lysergic acid diethylamide. *British Medical Journal,* 1966, **2,** 48–49.

Savage, C. The resolution and subsequent remobilization of resistance by LSD in psychotherapy. *Journal of Nervous and Mental Disease,* 1957, **125,** 434–437.

Savage, C. Psychedelic therapy. In J. M. Shlien (Ed.), *Research in psychotherapy.* Vol. 3. Washington, D. C.: American Psychological Association, 1968.

Savage, C., Fadiman, J., Mogar, R., & Allen, M. H. The effects of psychedelic (LSD) therapy on values, personality, and behavior. *International Journal of Neuropsychiatry,* 1966, **2,** 241–254.

Savage, C., Savage, E., Fadiman, J., & Harman, W. LSD: Therapeutic effects of the psychedelic experience. *Psychological Reports,* 1964, **14,** 111–120.

Schwartz, C. J. The complications of LSD: A review of the literature. *Journal of Nervous and Mental Disease,* 1968, **146,** 174–186.

Shagass, C., & Bittle, R. M. Therapeutic effects of LSD: A follow-up study. *Journal of Nervous and Mental Disease,* 1967, **144,** 471–478.

Shlien, J. M. (Ed.) *Research in psychotherapy.* Vol. 3. Washington, D. C.: American Psychological Association, 1968.

Smart, R. G., & Bateman, K. Unfavorable reactions to LSD: A review and analysis of the available case reports. *Canadian Medical Association Journal,* 1967, **97,** 1214–1221.

Tietz, W. Complications following ingestion of LSD in a lower class population. *California Medicine,* 1967, **107,** 396–398.

Unger, S. M. Mescaline, LSD, psilocybin and personality change. *Psychiatry,* 1963, **26,** 111–125.

Ungerleider, J. T., Fischer, D. D., & Fuller, M. The dangers of LSD: Analysis of seven months' experience in a university hospital's psychiatric service. *Journal of the American Medical Association,* 1966, **197,** 389–392.

Ungerleider, J. T., Fischer, D. D., Fuller, M., & Caldwell, A. The "bad trip"—The etiology of the adverse LSD reaction. *American Journal of Psychiatry,* 1968, **124,** 1483–1490.

Ungerleider, J. T., Fisher, D. D., Goldsmith, S. R., Fuller M., & Forgy, E. A statistical survey of adverse reactions to LSD in Los Angeles County. *American Journal of Psychiatry,* 1968, **125,** 352–357.

Welpton, D. F. Psychodynamics of chronic lysergic acid diethylamide use. *Journal of Nervous and Mental Disease,* 1968, **147,** 377–385.

(Received July 17, 1972)

Psychological Bulletin
1973, Vol. 79, No. 6, 352–361

THE AMSEL FRUSTRATION EFFECT:
INTERPRETATIONS AND RESEARCH[1]

JOHN W. SCULL[2]
University of Sussex

Experiments in the double runway and its analogs are reviewed and evaluated in terms of several theories that have been applied to this situation. Because of the complexity of the situation, a simple correspondence between the frustration effect (FE) and a hypothetical frustration response has not been convincingly demonstrated. Rather, a theoretical account of the FE must include possible effects of secondary reward, competing responses, generalization decrements, demotivation, and differential conditioning. Frustration theory is not clearly supported by the FE, but neither are there clear grounds for denying the role of frustration in producing it.

In 1952, Amsel and Roussel provided an apparently unambiguous demonstration of the motivating effects of frustration. They trained rats in two runways placed end to end so that the goal box of the first served as the start box for the second. After training with continuous reward in both goal boxes, reward in the first goal box was presented on an intermittent schedule. It was found that running speed in the second runway during partial reward was greater after the omission of reward in the first goal box (N trials) than it was after reward (R trials). Amsel and Roussel called this difference in speed the frustration effect (FE) and attributed it to increased motivation following omission of an expected reward. The present paper is a review of subsequent experiments on the FE and the theories that have been proposed to account for FE.

THEORIES OF THE FRUSTRATION EFFECT

Amsel's Frustration Theory

The principal theory of the FE has been the frustration theory elaborated by Amsel (1958, 1962). His position is that the frustration reaction includes a temporary increase in the strength of generalized drive. The antecedent condition of frustration is the omission of reward in the presence of an expectancy of reward, and the expectancy is identified with Spence's (1956) r_G–s_G mechanism. The amount of frustration should be directly related to the strength of r_G.

The success of Amsel's theory in explaining the FE has been evaluated in two ways: (*a*) tests of specific assertions in Amsel's theory and (*b*) experiments with control procedures or experimental manipulations derived from other interpretations of the FE. The first type of test is discussed in this section; experiments of the second type are discussed along with the theories from which they derive.

Magnitude of reward. Two basic experimental designs have been employed in studies of the effect of magnitude of reward on the FE. In between-groups experiments, different groups of rats have been trained with different magnitudes of intermittent reward in the first goal box. In within-subjects experiments, different stimuli in the first runway have been correlated with different magnitudes of intermittent reward in the first goal box.

Using the between-groups procedure, it has usually been found that the difference between R and N trials increases with magnitude of reward (Barrett, Peyser, & McHose, 1965; Daly, 1968; Krippner, Endsley, & Tacker, 1967). Seward, Pereboom, Butler, and Jones (1957) did not obtain this result. In addition, two studies (Carlson, 1968; Levy & Seward, 1969) found a statistically significant FE with a large incentive but failed to obtain the effect with a small incentive. In within-subjects experiments, it has been consistently found that the difference between R and N trials increases with magnitude of reward (McHose, 1969; McHose & Gavelek, 1969; McHose, Meyer, & Maxwell, 1969; Peckham & Amsel, 1964, 1967).

---

[1] This paper was prepared while the author was a National Research Council of Canada postdoctoral Fellow. The author is now at Cedar Lodge Residential Centre, Cobble Hill, British Columbia, Canada.

[2] Requests for reprints should be sent to John W. Scull, Cedar Lodge Residential Centre For Boys, Cobble Hill, British Columbia, Canada.

Frustration theory, however, makes the stronger prediction that absolute speed on N trials should increase with increasing magnitude of reward on R trials. Only one between-groups experiment (Krippner et al., 1967) obtained a difference on N trials. The others obtained differences on R trials only. In within-subjects experiments, Peckham and Amsel (1964) found greater speeds on N trials with increasing magnitude of reward; Peckham and Amsel (1967) and McHose et al. (1969) each found an N trial effect in one of their two experiments but not in the other. The other within-subjects experiments found effects on R trials only.

These inconsistent results on N trials are not explicable in terms of Amsel's theory, but Krippner et al. (1967) have made an additional assumption about the double runway that may resolve these differences. They have suggested that increasing magnitude of reward increases the secondary reward value of the first goal box. This produces a tendency for the subject to return to the first goal from the second runway. Speed in the second runway results from the combination of this tendency to retrace with the effects of frustration and reward in the second goal. On the basis of this three-factor interpretation, inconsistent results seem less surprising.

Another similar ad hoc account has been suggested by Wilton, Strongman, and Nerenberg (1969). They attributed their failure to obtain reliable FEs with pigeons to competing responses which they observed and suggested that failures to obtain increases in N-trial speed with increasing magnitude of reward may also be due to the effects of such competing responses. The danger inherent in this notion is that, without prior specification of the relative effects of motivation and competing responses in a particular experiment, almost any effect of frustration might be expected.

Stimuli in the first runway. Two studies have varied the similarity of the stimuli in the first runway to the stimuli in the first goal box (Amsel, Ernhart, & Galbrecht, 1961; Amsel & Hancock, 1957). Both found a larger FE when the two were more similar. As this operation was presumed to be a manipulation of the strength of r_G, these experiments are interpreted as supporting Amsel's theory. However, it is unclear why, in terms of Spence's (1956) theory, this manipulation should change the strength of r_G in the goal box itself.

Prior training. According to the frustration theory (Amsel, 1958), the size of the FE should vary as the expectancy of reward is manipulated by different types of training. Several different experimental designs have been employed in testing this prediction: (*a*) Partial reward has been given from the outset of training: The FE gradually increases as such training progresses (Amsel & Ward, 1965). (*b*) Partial reward has been introduced after a relatively small number of continuously rewarded trials: The FE has been found to be an increasing function of the number of prior rewards (Amsel & Ward, 1965; Hug, 1970a; Stimmel & Adams, 1969; Yelen, 1969). (*c*) Groups of rats have been given chronic nonreward in the first goal box and then shifted to partial reward: Under these conditions, the FE appears in the first few trials after the shift in reward percentage (Amsel & Prouty, 1959; Ison, Daly, & Glass, 1967; Ludvigson, 1967; McCain & McVean, 1967; Platt & Senkowski, 1970) and does not differ in size from the FE shown by a control group given continuous reward before the shift.

In the first two designs, acquisition of r_G is confounded with acquisition of the second instrumental response, while in the third design, the FE appears very quickly. None of these results, then, can be said to support frustration theory.

Yet another possible effect of training on the strength of r_G is provided by extinction. McHose (1963) shifted different groups of rats from continuous reward to either partial reward or extinction. He found that the FE disappeared as extinction progressed but remained unchanged in the partial reward group. In discriminative extinction it has been found that the FE, after nonrewarded trials to the negative stimulus, decreases as the discrimination develops (Amsel & Ward, 1965; Hug, 1970b). These experiments provide strong support for the idea that the FE depends on the strength of r_G, and the two experiments with partial reward control groups (Hug, 1970b; McHose, 1963) show that the effect is not simply due to some form of habituation to the effects of nonreward itself.

Variations in the frustrated response. In Amsel's theory, frustration is specifically defined as the disconfirmation of an expectancy, rather than as the thwarting of a response, an interpretation suggested by some other authors (e.g., Brown & Farber, 1951; Spence, 1960). On this basis, Amsel should predict little effect of the nature of the frustrated response on the FE, although this implication has not apparently always been appreciated by Amsel himself (Amsel et al., 1961). Variations in the length of the first runway of a double runway marginally affected the FE in one experiment (Amsel et al., 1961) and had no effect in another (Clifford & Schindelheim, 1968). Negative results have also been obtained with variations in the length of fixed-ratio bar-pressing schedules (Cole & VanFleet, 1970) and the effort required for the frustrated response (Carlson, 1968; Grusec & Bower, 1965). Correlating the partial reward schedule with the vigor of the first response also does not seem to affect the FE (Logan, 1968).

Incomplete incentive reduction. Beginning with an experiment by Bower (1962), the effects of incomplete incentive reductions in the double runway have been studied extensively.

Increasing the amount of reward magnitude reduction in the first goal box of the double runway has consistently been shown to increase running speed in the second runway (Barrett et al., 1965; Bower, 1962; Daly, 1968; Karabenick, 1969; McHose & Ludvigson, 1965; Patten & Myers, 1970). This result is analogous to effects with the FE produced by nonreward. However, it is necessary to show that the reduction in the amount of reward, rather than the absolute amount itself, is the variable affecting the FE. A control group that consistently receives the small reward in the first goal box, similar to the "Wagner" control group for reward-omission experiments to be discussed below, must be run. With this between-groups design, no evidence for an FE produced by incomplete reward magnitude reduction has been obtained (Barrett et al., 1965; Daly, 1968; Karabenick, 1969; McHose & Ludvigson, 1965).

However, Hall and Marr (1969) gave differential magnitude of reward training in the first runway of a double runway and then presented test trials on which the magnitudes of reward were reversed. They found that the large reward magnitude produced the same speed in the second runway, regardless of which stimulus had been present in the first runway; the small reward resulted in faster running after it was given in the large reward stimulus than when it was given in the small reward stimulus.

Most investigations of delay of reward in the double runway have found slower speeds in the second runway following delay of reward than following immediate reward (McHose, 1966a; Sgro, Glotfelty, & Moore, 1970; Sgro, Glotfelty, & Podlesni, 1969). However, Sgro (1969) found that speed increased when partial delay of reward was introduced and that the amount of increase was directly related to the amount of reward. He attributed the difference between his results and the other work to differences in the delay procedure used.

McHose (1966b, 1968) has obtained an FE both within subjects and in comparison with control groups for absolute incentive by simultaneously increasing delay of reward and reducing magnitude of reward. He found that the FE increased with increasing delay. In the first study, there was no effect of amount of magnitude reduction, but in the second, he found a negative relationship between the amount of magnitude reduction and the size of the FE. These demonstrations of FEs, with a combination of the two incentive variables, dictate caution when interpreting the failure of either variable alone to produce an FE in comparison with the appropriate control group.

Motivational nature of the frustration effect. In Amsel's theory, the FE is attributed to a temporary increase in drive, although the additional possibility of frustration-induced responses is recognized. There is considerable evidence that the FE does, in fact, represent the nonspecific effect on vigor implied by a motivational concept.

Several studies have obtained an FE measured as a difference in activity following R or N trials (Dunlap, Hughes, O'Brien, Lewis, & Dachowski, 1971; Gallup & Altomari, 1969; Gallup & Hare, 1969; Tacker & Way, 1968; Topping, Cole, Matamoros, & Linenberger, 1970). The FE was also demonstrated when force of bar pressing was measured immediately after intermittent reward of chain pulling (Levine & Loesch, 1967), and when aggression was measured after intermittent reward of running (Gallup, 1965). These experiments in which the frustrated response is different from the response used to measure frustration support the idea that the FE is a motivational effect.

One experiment (Schmeck & Bruning, 1968) has been directly addressed to the relationship of the FE to Spence's (1956) drive concept. In an apparatus consisting of a runway followed by either another runway or by a complex maze, these investigators found that frustrative nonreward in the first runway increased speed in the second runway but also increased the number of errors in the maze.

Evaluation of Amsel's theory. The success of Amsel's theory in dealing with these findings has clearly been rather limited. However, the ambiguity of the relationship of many of these results to frustration theory does not indicate that the theory is unfalsifiable. Rather, it indicates that the double runway is a much more complex situation than anticipated.

Learned Reactions to Frustration

Brown and Farber (1951) and Marx (1956) have suggested that, early in life, animals may learn to respond vigorously to frustration because vigorous, aggressive behavior has been frequently reinforced in frustrating situations in the subject's past. This differential reinforcement of vigorous responding when frustrated transfers to the double runway situation and produces the FE. Such a theory leads to all the same predictions about the FE as does the motivational account, excepting experiments involving prior reinforcement of vigor in another frustrating situation. The only experiment directly addressed to this proposal (Amsel & Penick, 1962) was inconclusive. All that can be said about the learned vigor interpretation, then, is that there is no evidence on the one point in which it differs from Amsel's theory.

Escape from Frustration

It has been suggested that the FE represents escape from frustration rather than increased drive after frustration (Leitenberg, 1965; Wagner, 1966). This proposal is the obverse of the secondary reward hypothesis proposed by Krippner et al. (1967) discussed above, and has about the same high face validity as an interpretation of the double-runway situation, where the subject does, in fact, escape from the frustrating situation. There have been, however, a number of demonstrations of the FE in operant situations where performing the second response did not remove the subject from the frustrating situation (Berger, 1969; Carlson, 1968; Cole & VanFleet, 1970; Davenport, Flaherty, & Dyrud, 1966; Platt & Senkowski, 1970).

Unreduced r_G

Bower (1962) has proposed that r_G is aroused by the stimuli in the first runway and goal box of the double runway and, subsequently, is reduced by reward in the first goal box. On nonrewarded trials, this r_G for the first reward is not reduced, but persists into the second runway, summates with the r_G for the second reward, and thereby invigorates responding. This theory makes the same predictions as frustration theory without assigning unique properties to frustrative nonreward.

Berger (1969) has deduced a similar account of the FE from Perkins' (1968) theory of preparatory responses. The preparatory response to the first reward persists on N trials, thereby increasing the value of the second reward. This leads to the prediction that the FE should be greater when both responses lead to the same type of reward than when they lead to different rewards. Berger himself (1969), however, found the same size FE regardless of whether the second response was motivated by food reward or shock avoidance; Levy and Seward (1969) found the same size FE with food in the first goal box and either food or water in the second; Amsel and Penick (1962) obtained an FE with food in the first goal box and a "social" animal in the second; and a number of studies have found an FE with presumably unrewarded activity as the indicant response (Dunlap et al., 1971; Gallup & Altomari, 1969; Gallup & Hare, 1969; Tacker & Way, 1968; Topping et al., 1970). There is, then, no support for the Berger-Perkins interpretation of the FE in the one area where it differs from the interpretations proposed by Amsel and Bower.

A similar account of the FE can be derived from the concepts of inertial tendency (Atkinson, 1964) or persisting tendency (Birch, 1968). These writers suggest that an r_G, once aroused by a stimulus, will persist for some time in the absence of stimulus support until a goal, like the one on which it is based, is reached.

The only difference between this view and the theories already discussed is that persisting tendencies have been shown to last for as long as 30 minutes (Birch, 1968), while the FE is marginally attenuated after 30 seconds' confinement in the first goal box (Amsel & Roussel, 1952), substantially attenuated after 90 seconds' confinement (MacKinnon & Amsel, 1964), and completely eliminated after 5 minutes' confinement (Robinson & Clayton, 1963).

The Demotivation Hypothesis

Seward et al. (1957) performed an FE experiment in which, in addition to partial reward in the first goal box, intermittent reward was given in the start box of the first runway before the beginning of the trial. They found that speeds in the first runway were depressed on prefeeding trials and suggested that both the prefeeding effect in the first runway and the FE in the second runway were due to demotivational effects of reward. They suggested several possible mechanisms for this demotivation.

As Amsel defined frustration in terms of disconfirmation of an expectancy rather than frustration of a response, there is no reason to assume, as Seward et al. (1957) apparently did, that intermittent prefeeding should not produce frustration. The effects of intermittent prefeeding obtained by Seward et al. and subsequent investigators (Kemble & Clayton, 1967; McCain & Power, 1966) appear to be another case of the absence of an effect on the FE of the task required in the first runway discussed earlier.

Maltzman (1952) reviewed the early literature on prefeeding and concluded that small amounts of prefeeding facilitate subsequent performances while larger amounts inhibit it. Daly (1968) found a similar effect with different magnitudes of reward in the first goal box of a double runway, and Dachowski (1967) found that different magnitudes of continuous reward in the first goal box of a double runway had no effect on running speed in the second runway. These inconsistent demotivation effects stand in sharp contrast to the surprising reliability of the FE. It seems unlikely that an effect which is elusive when studied directly can account entirely for a different phenomenon which is marked by the frequency and consistency with which it has been demonstrated.

The inadequacy of the demotivation hypothesis alone was most convincingly demonstrated in an experiment by Wagner (1959). In addition to groups of rats showing the FE, he ran a control group that was never rewarded in the first goal gox. The FE was assessed by comparing performance of the partial reward groups on N trials with performance of the control group, for which all trials were N trials but presumably not frustrating. He found a reliable between-subjects FE measured in this way which cannot be explained in terms of demotivation: Nonreward after a history of reward produced faster running than nonreward without the prior rewards. Since Wagner's (1959) original experiment, this "Wagner control" group has been used frequently. A few studies have failed to find the FE measured between-subjects in this way (Kemble & Clayton, 1967; Prytula & Ludvigson, 1967; McHose, 1963; Prytula & Braud, 1969), and Daly (1968) only obtained a reliable between-subjects effect in one of three pairs of groups, but most studies have obtained the between-subjects FE (Barret et al., 1965; Dunlap & Frates, 1970; Dunlap et al., 1971; Hamm, 1967; Ison et al., 1967; McCain & McVean, 1967; Robinson & Clayton, 1963; Schmeck & Bruning, 1968; Strongman & Wookey, 1969).

Generalization Decrement Hypothesis

In addition to possessing possible motivational properties, the events in the first goal box of a double runway may serve as part of the stimulus complex controlling the second response. The first few trials after any change in conditions, then, should be interpreted with caution (Amsel & Prouty, 1959). Further, on fixed-interval schedules in operant procedures, reward is an important cue controlling the fixed-interval "scallop" (e.g., Staddon, 1967). Omission of reward on these schedules, then, removes the principal cue for slow responding, producing an increase in response rate that may be entirely due to a generalization decrement.

Staddon (1970) has suggested that this analysis may not be specific to fixed-interval schedules but may be appropriate to FE experiments in general. Most of the responses used to measure the FE show a positive goal gradient, although the effect is not usually as marked as it is with fixed-interval schedules. After continuous reward, therefore, the aftereffects of reward should have a conditioned inhibitory effect on behavior through their association with slow responding early in the chain. Omitting reward produces a generalization decrement, which in this case results in an increase in response speed in the early part of the runway or response chain.

Demonstrations that reward can acquire control over subsequent responding (e.g., Calef, Meyer, Maxwell, McHewitt, & McHose, 1969; Staddon, 1970), however, do not imply that this must always be the case, and there are several FE experiments that seem to present difficulties for Staddon's position. It is not clear in this framework why the FE should persist with continued partial reward training (Hug, 1970b; McHose, 1963) but disappear with continued extinction (McHose, 1963) or discrimination training (Amsel & Ward, 1965; Hug, 1970b). The findings that shifts from chronic nonreward to intermittent reward produce an almost immediate FE (Amsel & Prouty, 1959; Ison, Daly, & Glass, 1967; Ludvigson, 1967; McCain & McVean, 1967; Platt & Senkowski, 1970) appear to be as difficult for Staddon's theory as they are for Amsel's. In terms of Staddon's theory, the sudden introduction of reward should have the opposite effect.

In order to explain any of these results, additional assumptions must be added to Staddon's theory. In view of the lack of explicit assumptions in Staddon's (1970) paper, it is impossible to properly evaluate this interpretation of the FE.

Differential Conditioning Interpretations

The double runway and its operant analogs can be viewed as a special case of differential conditioning in which the two stimuli are presented in closely successive pairs in a fixed order (Daly, 1968; McHose, 1970). That is, responses in each runway are rewarded on particular schedules, which may be different; and the two runways are discriminable on the basis of either order of presentation or physical differences. Thus, in addition to any effects attributed to aftereffects of the event in the first goal box, the contrast and generalization effects generally observed in differential conditioning (Black, 1968; Dunham, 1968; McHose, 1970) should also be observed in the double runway. Daly (1968) first elaborated this view in accounting for effects of magnitude of continuous reward in the double runway and used it, in combination with frustration theory, to account for her incentive-reduction results.

McHose (1970) has presented a more systematic account, arguing that the results of double runway experiments may be adequately attributed to (*a*) effects of differential conditioning and (*b*) demotivational properties of reward. He has further proposed a description of differential conditioning that predicts negative contrast or generalization effects in both positive and negative stimuli. The size of the effect is proportional to the difference between the incentives signalled by the two stimuli.

The within-subjects FE, according to Mc-

Hose, is entirely due to demotivational effects of reward, but the between-subjects FE obtained with the Wagner control group is due to both factors. Chronic nonreward in the first goal produces slower running in the second runway for the Wagner control group than does the continuous or partial reward given to the experimental group (contrast or generalization effects). Combination of these two partially independent effects, then, can potentially explain not only the between-subjects FE but also the reported failures to obtain this result.

There are, however, some weaknesses in his analysis. First, as has been pointed out, the demotivating effects of reward are not found as consistently as the FE. The second weakness lies in the failure to explain the immediate occurrence of the FE when chronic nonreward is changed to reward (McCain & McVean, 1967). In this experiment, the difference between N trials and R trials not only appeared almost immediately, but represented an increase in speed on N trials rather than a decrease on R trials. As differential conditioning is presumably an associative process like the conditioning of r_G in Amsel's (1958) theory or the development of inhibitory control in Staddon's (1970), this experiment presents the same problem for all three analyses.

It also seems that in experiments showing the effect of magnitude of reward on the FE, the analysis proposed by McHose has as inconsistent a record as frustration theory. McHose's analysis seems to predict that, if reward in the first goal box is larger than reward in the second, then N-trial speed should be inversely proportional to magnitude of first goal box reward. This result, the opposite of the one predicted by Amsel, has not been obtained (Barrett et al., 1965; Krippner et al., 1967; McHose & Ludvigson, 1965).

The strength of this analysis lies in its success at explaining most of the experiments on incomplete reduction in magnitude of reward better than does frustration theory. In addition, direct tests of predictions from the theory in differential conditioning studies with short intertrial intervals on some trials have generally confirmed the theory (Calef et al., 1969; Fox, Calef, Gavelek, & McHose, 1970; McHewitt, Calef, Maxwell, Meyer, & McHose, 1969). However, it is not clear that this interpretation is superior to Daly's (1968) combination of differential conditioning and frustration effects rather than differential conditioning and demotivation effects.

CONCLUSIONS

While early experiments on the FE were taken as strong evidence for the existence of frustration drive (e.g., Amsel, 1958, 1962; Mowrer, 1960; Spence, 1960; Yates, 1962), subsequent research and theorizing has shown that the double runway is such a complex situation that this basic conclusion may have been unwarranted. The numerous accounts of the double runway may be classified according to their success in accounting for the data, although all have been important in suggesting control procedures and new directions for investigation.

- Theories of persisting tendencies (Atkinson, 1964; Berger, 1969; Birch, 1968), escape from frustration (Leitenberg, 1965; Wagner, 1966), and generalization decrements (Staddon, 1970) cannot alone account for a number of results.
- Several theories are isomorphic to Amsel's (1958) frustration theory in the double runway (Bower, 1962; Brown & Farber, 1951; Marx, 1956), and will share the fate of that analysis. Amsel's theory itself, like the demotivation hypothesis (Seward et al., 1957), cannot account for all the results without the addition of other assumptions.
- Variations of the demotivation hypothesis (McHose, 1970) or frustration theory (Daly, 1968; Krippner et al., 1967) seem most consonant with the existing literature. However, the complexity of these proposals makes a choice among them impossible at the present time.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin,* 1958, **55**, 102–119.

AMSEL, A. Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychological Review,* 1962, **69**, 306–328.

Amsel, A., Ernhart, C. B., & Galbrecht, C. R. Magnitude of frustration effect and strength of antedating goal factors. *Psychological Reports*, 1961, **8**, 183–186.

Amsel, A., & Hancock, W. Motivational properties of frustration: III. Relation of frustration effect to antedating goal factors. *Journal of Experimental Psychology*, 1957, **53**, 126–131.

Amsel, A., & Penick, E. C. The influence of early experience in the frustration effect. *Journal of Experimental Psychology*, 1962, **63**, 167–176.

Amsel, A., & Prouty, D. L. Frustrative factors in selective learning with reward and nonreward as discriminanda. *Journal of Experimental Psychology*, 1959, **57**, 224–230.

Amsel, A., & Roussel, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *Journal of Experimental Psychology*, 1952, **43**, 363–368.

Amsel, A., & Ward, J. S. Frustration and persistence; Resistance to discrimination following prior experience with the discriminanda. *Psychological Monographs*, 1965, **79**(4, Whole No. 597).

Atkinson, J. W. *An introduction to motivation.* Princeton: Van Nostrand, 1964.

Barrett, R. J., Peyser, C. S., & McHose, J. H. Effects of complete and incomplete reward reduction on a subsequent response. *Psychonomic Science*, 1965, **3**, 277–278.

Berger, D. F. Alternative interpretations of the frustration effect. *Journal of Experimental Psychology*, 1969, **81**, 475–483.

Birch, D. Shift in activity and the concept of persisting tendency. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* Vol. 2. New York: Academic Press, 1968.

Black, R. W. Shifts in magnitude of reward and contrast effects in instrumental and selective learning: A reinterpretation. *Psychological Review*, 1968, **75**, 114–126.

Bower, G. H. The influence of graded reductions in reward and prior frustrating events upon the magnitude of the frustration effect. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 582–587.

Brown, J. S., & Farber, I. E. Emotions conceptualized as intervening variables with suggestions toward a theory of frustration. *Psychological Bulletin*, 1951, **48**, 465–495.

Calef, R. S., Meyer, P. A., Maxwell, F. R., McHewitt, E. R., & McHose, J. H. Synthesis of double alley and discrimination phenomena: Cue and magnitude specific properties of G_1 reward. *Psychonomic Science*, 1969, **17**, 39–41.

Carlson, J. G. Frustrative nonreinforcement of operant responding: Magnitude of reinforcement and response force effects. *Psychonomic Science*, 1968, **11**, 307–308.

Clifford, T., & Schindelheim, R. H. The frustration effect as a function of runway length. *Psychonomic Science*, 1968, **10**, 109–110.

Cole, M., & VanFleet, F. M. The frustration effect as a function of interoceptive and exteroceptive cues in a Skinner box analogue of the double runway. *Psychonomic Science*, 1970, **20**, 33–35.

Dachowski, L. Reward size effects in the double runway. *Psychonomic Science*, 1967, **9**, 157–158.

Daly, H. B. Excitatory and inhibitory effects of complete and incomplete reward reduction in the double runway. *Journal of Experimental Psychology*, 1968, **76**, 430–438.

Davenport, J. W., Flaherty, C. F., & Dyrud, J. P. Temporal persistence of frustration effects in monkeys and rats. *Psychonomic Science*, 1966, **6**, 411–412.

Dunham, P. J. Contrasted conditions of reinforcement: A selective critique. *Psychological Bulletin*, 1968, **69**, 295–315.

Dunlap, W. P., & Frates, S. B. Influence of deprivation on the frustration effect. *Psychonomic Science*, 1970, **21**, 1–2.

Dunlap, W. P., Hughes, L. F., O'Brien, T. J., Lewis, J. H., & Dachowski, L. Goalbox activity as a measure of frustration in a single runway. *Psychonomic Science*, 1971, **23**, 327–328.

Fox, P. A., Calef, R. S., Gavelek, J. R., & McHose, J. H. Synthesis of differential conditioning and double alley data: Performance to S+ as a function of intertrial interval and antedating reward events. *Psychonomic Science*, 1970, **18**, 141–143.

Gallup, G. G., Jr. Aggression in rats as a function of frustrative nonreward in a straight alley. *Psychonomic Science*, 1965, **3**, 99–100.

Gallup, G. G., Jr., & Altomari, T. S. Activity as a postsituation measure of frustrative nonreward. *Journal of Comparative and Physiological Psychology*, 1969, **68**, 382–384.

Gallup, G. G. Jr., & Hare, G. K. Activity following partially reinforced trials: Evidence for a residual frustration effect due to conditioned frustration. *Psychonomic Science*, 1969, **16**, 41–42.

Grusec, T., & Bower, G. Response effort and the frustration hypothesis. *Journal of Comparative and Physiological Psychology*, 1965, **60**, 128–130.

Hall, L. A., & Marr, J. N. Incomplete reduction of reward and the frustration effect with hunger constant. *Journal of Experimental Psychology*, 1969, **80**, 493–497.

Hamm, H. D. Perseveration and summation of the frustration effect. *Journal of Experimental Psychology*, 1967, **73**, 196–203.

Hug, J. J. Frustration effects after varied numbers of partial and continuous reinforcements: Incentive differences as a function of reinforcement percentage. *Psychonomic Science*, 1970, **21**, 57–59. (a)

Hug, J. J. Frustration effect after development of patterned responding to single-alternation reinforcement. *Psychonomic Science*, 1970, **21**, 61–62. (b)

Ison, J. R., Daly, H. B., & Glass, D. H. Amobarbital sodium and the effects of reward and nonreward in the Amsel double runway. *Psychological Reports*, 1967, **20**, 491–496.

Karabenick, S. A. Effect of reward increase and reduction in the double runway. *Journal of Experimental Psychology*, 1969, **82**, 79–87.

Kemble, E. D., & Clayton, K. N. Prefeeding and the apparent frustration effect. *Psychonomic Science*, 1967, **9**, 491–492.

Krippner, R. A., Endsley, R. C., & Tacker, R. S. Magnitude of G_1 reward and the frustration effect in a between subjects design. *Psychonomic Science*, 1967, **9**, 385–386.

Leitenberg, H. Is time-out from positive reinforcement an aversive event? *Psychological Bulletin*, 1965, **64**, 428–441.

Levine, G., & Loesch, R. Generality of response intensity following non-reinforcement. *Journal of Experimental Psychology*, 1967, **75** 97–102.

Levy, N., & Seward, J. P. Frustration and homogeneity of rewards in the double runway. *Journal of Experimental Psychology*, 1969, **81**, 460–463.

Logan, F. A. Frustration effect following correlated nonreinforcement. *Journal of Experimental Psychology*, 1968, **78**, 396–400.

Ludvigson, H. W. A preliminary investigation of the effects of sodium amytal, prior reward in G1, and activity level on the FE. *Psychonomic Science*, 1967, **8**, 115–116.

MacKinnon, J. R., & Amsel, A. Magnitude of the frustration effect as a function of confinement and detention in the frustrating situation. *Journal of Experimental Psychology*, 1964, **67**, 468–474.

Maltzman, I. The process need. *Psychological Review*, 1952, **59**, 40–48.

Marx, M. H. Some relations between frustration and drive. *Nebraska Symposium on Motivation*, 1956, **4**, 92–130.

McCain, G., & Power, R. Extinction as a function of reinforcement conditions in the start box. *Psychonomic Science*, 1966, **5**, 193–194.

McCain, G., & McVean, G. Effects of prior reinforcement or nonreinforcement on later performance in a double alley. *Journal of Experimental Psychology*, 1967, **73**, 620–627.

McHewitt, E. R., Calef, R. S., Maxwell, F. R., Jr., Meyer, P. A., & McHose, J. H. Synthesis of double alley and discrimination phenomena: Apparent positive S+ contrast in differential conditioning. *Psychonomic Science*, 1969, **16**, 137–139.

McHose, J. H. Effect of continued nonreinforcement on the frustration effect. *Journal of Experimental Psychology*, 1963, **65**, 444–450.

McHose, J. H. Delay increase and subsequent responding. *Psychonomic Science*, 1966, **5**, 213–214. (a)

McHose, J. H. Incentive reduction: Simultaneous delay increase and magnitude reduction and subsequent responding. *Psychonomic Science*, 1966, **5**, 215–216. (b)

McHose, J. H. Incentive reduction: Varied simultaneous reductions and subsequent responding. *Psychonomic Science*, 1968, **11**, 313–314.

McHose, J. H. Role of frustration in the development of relative and absolute S− discrimination contrast effects. *Journal of Experimental Psychology*, 1969, **81**, 256–260.

McHose, J. H. Relative reinforcement effects: S_1/S_2 and S_1/S_1 paradigms in instrumental conditioning. *Psychological Review*, 1970, **77**, 135–146.

McHose, J. H., & Gavelek, J. R. The frustration effect as a function of training magnitude: Within- and between-*Ss* designs. *Psychonomic Science*, 1969, **17**, 261–262.

McHose, J. H., & Ludvigson, H. W. Role of reward magnitude and incomplete reduction of reward magnitude in the frustration effect. *Journal of Experimental Psychology*, 1965, **70**, 490–495.

McHose, J. H., Meyer, P. A., & Maxwell, F. R., Jr. Frustration effect as a function of training magnitude in a within-*S* design. *Psychonomic Science*, 1969, **14**, 137–138.

Mowrer, O. H. *Learning theory and behavior*. New York: Wiley, 1960.

Patten, R. L., & Myers, D. B. Number of training trials and frustration effects of nonzero reward reduction in the double runway. *Psychonomic Science*, 1970, **18**, 291–292.

Peckham, R. H., Jr., & Amsel, A. Magnitude of reward and the frustration effect in a within-subjects design. *Psychonomic Science*, 1964, **1**, 285–286.

Peckham, R. H., Jr., & Amsel, A. Within-subject demonstration of a relationship between frustration and magnitude of reward in a differential magnitude of reward discrimination. *Journal of Experimental Psychology*, 1967, **73**, 187–195.

Perkins, C. C., Jr. An analysis of the concept of reinforcement. *Psychological Review*, 1968, **75**, 155–172.

Platt, J. R., & Senkowski, P. C. Effects of discrete-trials reinforcement frequency and changes in reinforcement frequency on preceding and subsequent fixed-ratio performance. *Journal of Experimental Psychology*, 1970, **85**, 95–104.

Prytula, R. E., & Braud, W. G. Absolute sucrose incentive reduction: Frustration or demotivation? *Psychonomic Science*, 1969, **15**, 249–250.

Robinson, A. W., & Clayton, K. N. Effect of duration of confinement in a nonbaited goal box on the "apparent frustration effect." *Journal of Experimental Psychology*, 1963, **66**, 613–614.

Schmeck, R. R., & Bruning, J. L. Task difficulty and the frustration effect. *Journal of Experimental Psychology*, 1968, **78**, 516–520.

Seward, J. P., Pereboom, A. C., Butler, B., & Jones, R. B. The role of prefeeding in an apparent frustration effect. *Journal of Experimental Psychology*, 1957, **54**, 445–450.

Sgro, J. A. Complete removal and delay of sucrose reward in the double alleyway. *Journal of Comparative and Physiological Psychology*, 1969, **69**, 442–447.

Sgro, J. A., Glotfelty, R. A. & Moore, B. D. Delay of reward in the double alleyway: A within-subjects versus between-groups comparison. *Journal of Experimental Psychology*, 1970, **84**, 82–87.

Sgro, J. A., Glotfelty, R. A., & Podlesni, J. A. Contrast effects and delay of reward in the double alleyway. *Psychonomic Science*, 1969, **16**, 29–31.

SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale University Press, 1956.

SPENCE, K. W. *Behavior theory and learning*. Englewood Cliffs: Prentice-Hall, 1960.

STADDON, J. E. R. Attention and temporal discrimination: Factors controlling responding under a cyclic-interval schedule. *Journal of the Experimental Analysis of Behavior,* 1967, **10**, 349–359.

STADDON, J. E. R. Temporal effects of reinforcement: A negative "frustration" effect. *Learning and Motivation,* 1970, **1**, 227–247.

STIMMEL, D. T., & ADAMS, P. C. The magnitude of the frustration effect as a function of the number of previously reinforced trials. *Psychonomic Science,* 1969, **16**, 31–32.

STRONGMAN, K. T., & WOOKEY, P. E. Frustrative reward. *Journal of Experimental Psychology,* 1969, **82**, 183–184.

TACKER, R. S., & WAY, J. Motivational properties of nonreward. *Psychonomic Science,* 1968, **10**, 103–104

TOPPING, J. F., COLE, J. M., MATAMOROS, M. M., & LINENBERGER, C. H. Motivational effects of differential magnitudes of reinforcement. *Psychonomic Science,* 1970, **19**, 25–26.

WAGNER, A. R. The role of reinforcement and nonreinforcement in an "apparent frustration effect." *Journal of Experimental Psychology,* 1959, **57**, 130–136.

WAGNER, A. R. Frustration and punishment. In R. N. Haber (Ed.), *Current research in motivation.* New York: Holt, Rinehart & Winston, 1966.

WILTON, R. N., STRONGMAN, K. T., & NERENBERG, A. Some effects of frustration in a free responding operant situation. *Quarterly Journal of Experimental Psychology,* 1969, **21**, 367–380.

YATES, A. J. *Frustration and conflict.* New York: Wiley, 1962.

YELEN, D. Magnitude of the frustration effect and number of training trials. *Psychonomic Science,* 1969, **15**, 137–138.

(Received July 22, 1971)

Psychological Bulletin
1973, Vol. 79, No. 6, 362–372

NEGLECTED DEVELOPMENTAL VARIABLES IN THE STUDY OF SPECIES IDENTIFICATION IN BIRDS[1]

GILBERT GOTTLIEB[2]

North Carolina Department of Mental Health
Raleigh, North Carolina

A question of the highest developmental and perceptual significance launched the experimental analysis of imprinting: How do animals come to identify members of their own species in the course of ontogeny? As laboratory research on imprinting in birds has become more and more removed from that primary question, it threatens to continue to neglect certain fundamental natural history and neuroembryological facts which are basic to the study of the development of species-specific perception in birds and perhaps other animals. These problems are discussed in the context of an empirical review of the development of the auditory and visual components of species identification in wild and domestic ducklings.

In the present review the initial response of young ducklings to certain patterns of auditory and visual stimulation is examined in order to determine their preferences and especially to find out how selective the "naïve" hatchling's perception of species-specific stimulation is prior to entry into the imprinting situation. The question is whether young ducklings are uniquely sensitive to patterns of auditory and visual stimulation provided by adult members of their own species in advance of actual contact or exposure to adults. To the extent that "parentally naïve" ducklings are selectively responsive to certain aspects of species-typical parental stimulation, an extremely interesting problem is posed for prior ontogenetic analysis, namely, how in the course of previous individual development does this selective process of perception come about? In most studies of imprinting, the "naïve" or preexperimental sensory and perceptual preferences of the to-be-imprinted animals are taken for granted or ignored. The purposes of the present review are (*a*) to place these particular sensory and perceptual phenomena in the context of the development of species identification and (*b*) to discuss their more general significance for the study of the development of sensory processes and perception.

OBSERVATIONS IN NATURE

In nature, ducks and certain other highly precocial bird species take leave of the nest soon (within from 1 to 2 days) after the young have hatched. In initiating the departure from a ground nest, the duck hen vocalizes recurrently and slowly walks away from her brood, thus presenting them with a certain pattern of auditory and visual stimulation which differs from species to species.[3] The brood quickly follows the hen and shortly thereafter they begin to feed for the first time. Prior to the departure from the nest,

---

[1] The author is grateful to Jo Ann Winfree and Mary Catharine Vick for their assistance. Financial support was provided by the North Carolina Department of Mental Health and by United States Public Health Service Research Grant HD-00878 from the National Institute of Child Health and Human Development.

[2] Requests for reprints should be sent to Gilbert Gottlieb, Dorothea Dix Hospital, Raleigh, North Carolina 27611.

[3] The megapodes (mallee-fowl, brush-turkeys) present a striking exception to the generalized description presented here. The eggs of these species are not incubated by the parent but are incubated by the heat arising from volcanic ash or the compost heap in which they are buried. Upon hatching, the young flee from the nest-site. If they happen to encounter their parents in the vicinity, they pay them no mind. Although their parents show great concern (behaviorally) about their eggs, they appear to have no interest in their offspring. On the second or third day after hatching, the young are sufficiently precocious to fly up into brush or trees where they roost, apparently in solitude (Frith, 1962). The development of species identification in this group of birds, most of which inhabit the region of Australia, presents an intriguing analytic problem. A similar problem is presented by the parasitic species (cuckoos, cowbirds, et al.) which lay their eggs in the nests of other species of birds. It would be especially interesting to test the possibility that self-stimulation plays an important role in the ontogeny of species-specific perception in the megapodes and the parasitic egg-laying species.

the hen offers tactile, thermal, auditory, and occasionally, visual stimulation to her brood (Bjärvall, 1968; Collias & Collias, 1956; Gottlieb, 1963, 1965). During the nest departure she presents them only with auditory and visual stimulation.[4] It is of obvious adaptive advantage for the young to promptly approach, follow, and stay near the hen when she quits the nest. It is in the postnesting period that the young duckling's thermoregulatory ability matures and their acquaintance with their postnatal environment develops.

TABLE 1

BURST, RATE, AND DOMINANT FREQUENCY CHARACTERISTICS OF MATERNAL CALLS USED IN BEHAVORIAL EXPERIMENTS WITH DUCKLINGS

Call	Note		Dominant frequency (in Hz.)
	Single burst	Average rate per second	
Mallard hen	9	4.1	780, 1,125, 1,600, 2,300
Chicken hen	7	2.5	775
Wood duck hen	13	7.4	1,300
Pintail hen	9	5.6	440
Mandarin hen	9	9.0	1,325

Note. For experimental purposes, the same burst of any given call is repeated at approximately 3-second intervals for the duration of the tape. The rate of each call was determined by dividing the number of notes by the total time from the beginning of the first note to the end of the last note in the burst. Since the notes do not occur at equal intervals within the burst, the rate presented for each call is an average for the entire burst.

AUDIOVISUAL STIMULATION

Eric Fabricius (1951), expanding on the more casual observations of Lorenz (1935, 1937), was among the first to observe that the combined presence of rather general auditory and visual stimulation substantially increased the liklihood of eliciting a following response from young Tufted, Eider, Mallard, and Shoveler ducklings, and his finding has since been confirmed a number of times and in several different species (Redhead ducklings: Collias & Collias, 1956; Mallard ducklings: Boyd & Fabricius, 1965; Peking ducklings—highly domesticated form of Mallard: Gottlieb, 1963; Mallard & Wood ducklings: Gottlieb, 1968; Bobwhite quail chicks: Porter & Stettner, 1968; domestic chicks: Smith & Bird, 1963). These findings articulate well with the situation in nature, in which the following response of young precocial birds in many duck species is initially elicited by the combined presence of auditory and visual stimuli emanating from the maternal parent as she leaves the nest. These findings also make good neurological sense and accord with the idea of intersensory facilitation or summation as potentiating behavioral responsiveness.

A very important question left unanswered by intersensory facilitation, however, concerns the requisite specificity of the audiovisual perceptual aggregate: Are the auditory and visual perceptual properties of the parent bird in any way unique in initiating the following response of the young of her species? To answer this question, one must test incubator-hatched birds of various species with the maternal audiovisual configuration of their own species and other species. Consequently, the author recorded the vocalizations (Table 1) of various duck species in nature in the act of leading their young from the nest (Gottlieb, 1963, 1965, 1971a) and had ducks of the various species stuffed and fitted with an internal speaker so the calls could be played through them.[5] The next step was to present these moving and calling visual replicas individually to young ducklings which had been incubated and hatched in the laboratory. The author

---

[4] In hole-nesting species such as the Wood duck (*Aix sponsa*), the hen initially presents only auditory stimulation to her young during the nest departure Gottlieb, 1963, 1965).

[5] For a precise study of the parameters of audiovisual perception it is not a good practice to have the calls emanate from speakers inside the visual replicas because the internal characteristics of the replica (and the way the speaker is mounted) can cause acoustic distortions which can influence the bird's response. Most recently, therefore, we have used a very high quality midrange dome radiator speaker (Acoustic Research, Cambridge, Massachusetts), which has a virtually flat frequency response from 200–16,000 hertz mounted *outside* of the visual replica. This is a particularly important methodological point in simultaneous audiovisual choice tests where the same call is produced from (behind) both replicas rather than from internal speakers. In the latter case, we have observed a response bias which was most likely attributable to an alteration of frequency (pitch) or timbre produced by one of the speakers in one of the replicas. In any event, that particular (unpublished) result could not be reproduced when the undistorted call emanated from the wall-mounted speakers, whereas our other results were replicable under the improved acoustic conditions.

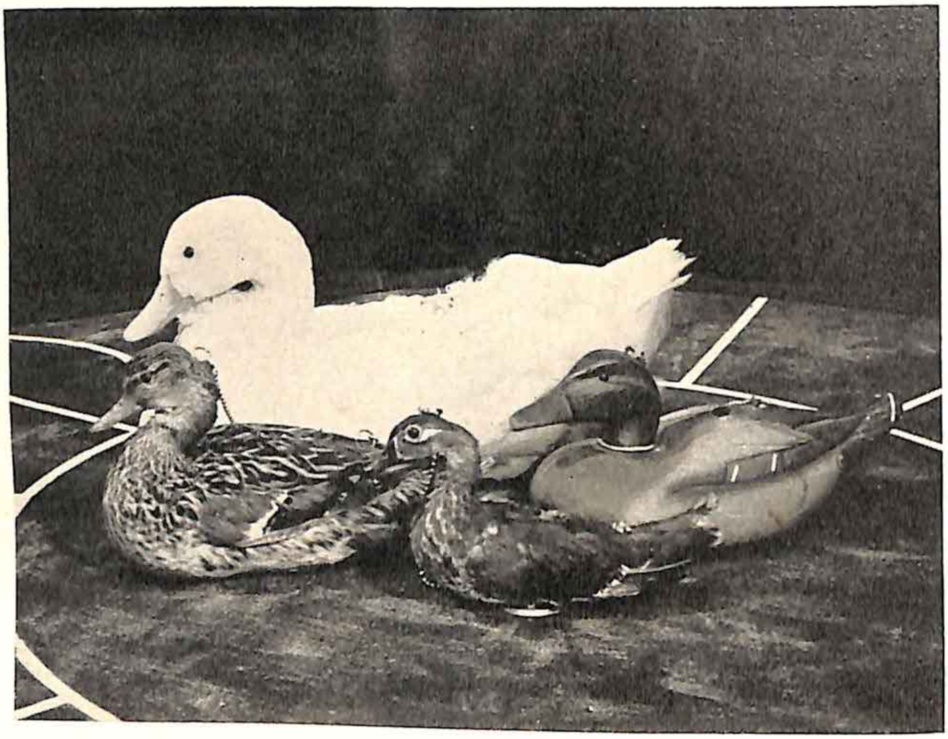


FIG. 1. Visual replicas used in the present experiments: Peking hen (at rear), Mallard hen (left), Mallard drake (right), and Wood duck hen (center front).

began by assessing the ducklings' following response to the maternal audiovisual configuration of their own species in a 20-minute test during the first 2 days after hatching. In these experiments and the other ones alluded to in this review, all the birds were hatched in the laboratory and maintained in individual visual isolation in brooders in light- and sound-treated rooms designed for these specific purposes. Unless noted to the contrary, no attempt was made to manipulate the birds' experiences prior to testing, so these results pertain to birds which had heard but not seen their siblings (conspecifics) and had neither heard nor seen maternal stimuli prior to testing. Since the amount of prior exposure to auditory stimulation from self and siblings plays a role in the response to the maternal call at testing (Gottlieb, 1966, 1971a), an explicit effort was made to maintain the birds in groups of no less than 20 each in the incubator and the brooder prior to the test. Each bird was tested individually and rather rigorous criteria were applied in measuring their approach or following response (described in detail in Gottlieb, 1971a, Ch. 2). The visual replicas are depicted in Figure 1. To appreciate the design of the experiments involving Mallard ducklings and their domestic derivatives (Peking ducklings), the reader should know that these animals are of the same species (*Anas platyrhynchos*) and that artificial selection has changed the visual characteristics of the hen (color, size) but not her vocal repertoire. Thus, in the first experiment below, the Mallard ducklings are tested with the Mallard maternal call emanating from the Mallard hen and the Peking ducklings are tested with the same call issuing from the Peking hen (see Figure 1). The Wood ducklings (*Aix sponsa*) are tested with the Wood duck maternal call emanating from the Wood duck hen.

As can be seen in Figure 2, the species-specific maternal audiovisual configurations elicited following in 84% of the Wood ducklings, 95% of the Mallard ducklings, and 92% of the Peking ducklings. When the visual component is deleted and the birds are tested to the maternal call of their species emanating from a nonvisible, moving sound source, these figures drop to 59%, 50%, and 52%, respec-

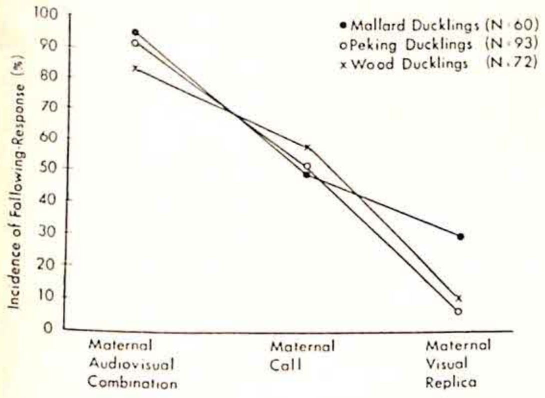


FIG. 2. The percentage of Mallard, Peking, and Wood ducklings which followed their respective species maternal call and a visual replica of a hen of their species, separately and in combination. (Data taken from Gottlieb, 1971a.)

tively. When the auditory component is deleted and the birds are tested to the silent, moving, visual replica of a hen of their own species, the figures drop even further to 11%, 30%, and 7%, respectively. Thus, the species-specific maternal audiovisual configuration elicits a higher degree of following than the purely auditory component, and the auditory component is more effective than the purely visual component.

To examine the question of the possible uniqueness of the species' maternal audiovisual configuration, the Wood duck maternal call was substituted for the Mallard call, and Mallard and Peking ducklings were tested with the Wood duck call emanating from their respective maternal visual replicas (Mallard hen and Peking hen). In this case, their following response dropped to 60% (Mallards) and 70% (Pekings) from the 90% level when the Mallard call issued from these replicas (see Figure 3). In order to assess the effect of changing the maternal visual component as well as the maternal auditory component, Mallard and Peking ducklings were tested for their following response when the Wood duck call emanated from a Mallard drake and a Wood duck hen. In this instance, 69% of the Mallards and 51% of the Pekings followed the drake, and 50% of the Mallards and 44% of the Pekings followed the Wood duck hen (Figure 3).

So far, then, the results indicate that the species-specific maternal audiovisual configuration is the single most effective object for the elicitation of the following response. When a substitution was made in the auditory component, the response declined significantly and when substitutions were made in both the auditory and visual components, the Pekings' response declined somewhat further and the decline in the Mallards' response remained about the same.

In the final experiment on this topic, only the visual component was changed: The Mallard call was made to emanate from the Mallard drake and the Wood duck hen. Under these conditions, 83% of the Mallards and 90% of the Pekings followed the Mallard drake and 100% of the Mallards and 100% of the Pekings followed the Wood duck hen. Thus, when the visual component is changed and the species-specific auditory component is retained, the following response of Peking and Mallard ducklings climbs back to the previous high level obtained when the species' maternal call emanates from the maternal replica. (The outcome of these last two experiments is summarized in Figure 3.)

These results indicate two things. First, the species-specific maternal audiovisual configuration is not uniquely effective in eliciting the following response and, second, the attractiveness of the species' maternal call seems to override (or compensate for) changes in the visual component. In essence, there is little or no decline in the following response of Mallard and Peking ducklings when the

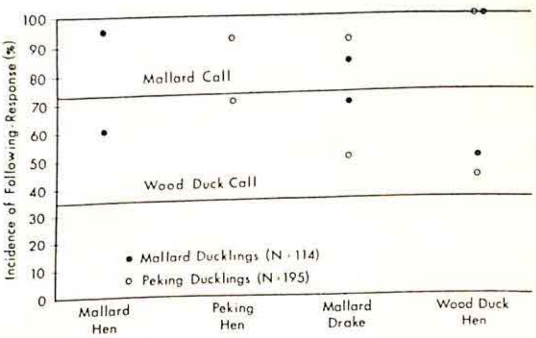


FIG. 3. The percentage of Mallard and Peking ducklings which followed various visual replicas when the replicas emitted either the Mallard maternal call or the Wood duck maternal call. (In all cases the presence of the Mallard call instigated a higher degree of following than did the Wood duck call. Data taken from Gottlieb, 1971a.)

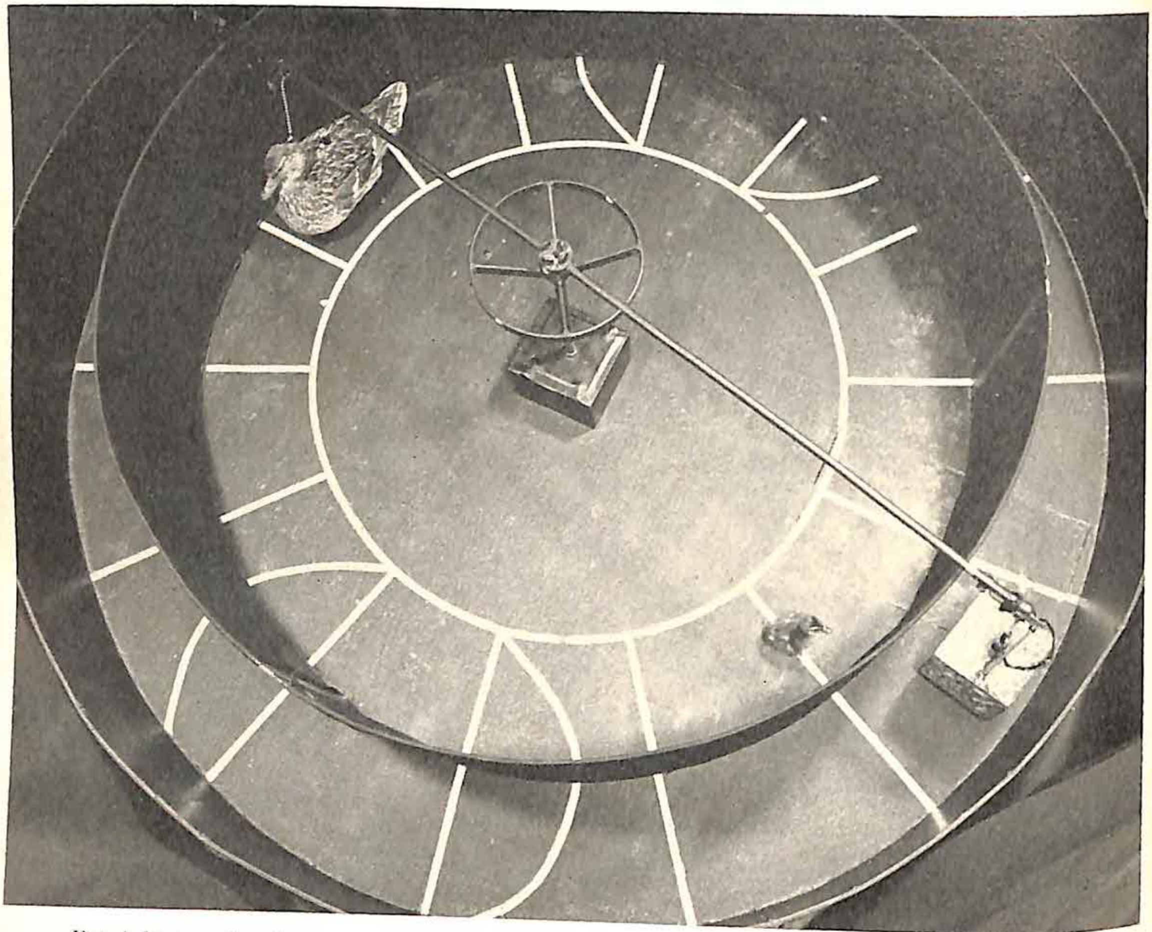


FIG. 4. Maternal auditory versus maternal visual choice test. (In this test the replica of a silent, moving hen is placed in opposition to her call emanating from a moving, nonvisible sound source—a speaker mounted inside a plywood box behind an opaque curtain. A Mallard duckling is shown following the source of the call in the above photograph.)

Mallard call accompanies any of the visual replicas, but there is a significant decline in responsiveness when it is replaced by another call, even when that call issues from the Mallard hen or the Peking hen. These results suggest that, in the initial stage of postnatal development in ducklings, species identification is based primarily on the auditory component of the maternal audiovisual-stimulating complex. They also suggest that the species-specific auditory component may be unique in its effectiveness in eliciting the response of young ducklings. These two different propositions are examined further in the next two sections.

TABLE 2

INCIDENCE OF APPROACH RESPONSE OF PEKING AND MALLARD DUCKLINGS TO MATERNAL CALLS OF VARIOUS SPECIES

Ducklings	Maternal calls				
	Mallard	Pintail	Wood duck	Mandarin	Chicken
Pekings ($n = 125$)	76%	80%	60%	28%	56%
Mallards ($n = 101$)	81%	80%	40%	25%	75%

Note. Each bird tested only once with each call; data taken from Gottlieb, 1971a.

AUDITORY VERSUS VISUAL STIMULATION

To further document the suggestion that species identification is based primarily on the auditory aspect of the maternal audiovisual configuration, the two components (moving) were presented separately but simultaneously to Wood ducklings and to Mallard ducklings

TABLE 3

Preference of Peking and Mallard Ducklings in Simultaneous Auditory Approach Tests

Choice test calls	*n*	Responded (*n*)	Preference (*n*)		
			Mallard	Other	Both
Pekings					
Mallard versus Pintail	21	19	19	0	0
Mallard versus Wood duck	44	27	24	3	0
Mallard versus chicken	30	24	24	0	0
Mallard					
Mallard versus Pintail	20	19	18	1	0
Mallard versus Wood duck	22	19	18	1	0

Note. Data taken from Gottlieb, 1971a.

in a choice test situation depicted in Figure 4. Although no behavioral evidence of conflict was observed, only about half the birds responded, and all of them followed the call. (The response figure—53% for the Wood ducklings and 45% for the Mallards—is close to the level of response addressed to the maternal call presented by itself under the test conditions [Figure 2].) These results support the notion that species identification is founded primarily on the auditory stimulation provided by the parent bird.

Auditory Stimulation

The next question concerns the uniqueness of the species' maternal call in the development of species identification: Are incubator-hatched birds uniquely sensitive to the maternal call of their own species? To answer this question, Peking and Mallard ducklings were tested for their response to a variety of maternal calls in a 4-minute approach test, each bird being tested only once to one of the calls. As shown in Table 2, the birds are equally responsive to several of the calls in the series, and there is thus no evidence of the unique attractiveness of their own species call under these test conditions. When the birds are given *simultaneous* auditory choice tests, however, pitting the maternal call of their species against the maternal calls of the other species to which they are very responsive in the single-call test, the Pekings and Mallards do show a strong and consistent preference for the Mallard maternal call in every case (Table 3).

Thus, although the birds approach a rather wide range of acoustic patterns presented singly, they demonstrate a unique *preference* for the particular acoustic pattern provided by their own species' maternal call when it is placed in direct (simultaneous) comparison with the maternal calls of other species. The unique sensitivity or preference which the birds have for the maternal call of their own species is revealed only in simultaneous choice tests.

Developmental Background of the Auditory and Visual Aspects of Species Identification: Problems and Considerations for Future Research

Upon considering the ducklings' response to the auditory and visual components of a hen of their species during the first few days after hatching, a number of different lines of evidence point to the prepotency of the auditory component over the visual component during this early period. In this instance, "prepotency" means that on any measure the maternal call plays a more powerful energizing and directive influence than a visual replica of the hen. The behavioral evidence for the prepotency of the auditory component over the visual component conforms rather well to what is known about the neuroembryological development of the sensory systems, as well as the embryo's and hatchling's sensory and perceptual experience prior to its entrance into the testing situation. The latter very possibly has a most significant bearing on the present results. Before discussing those important factors, let us briefly review some other behavioral evidence for the primacy of the auditory perceptual system over the visual system early in avian development.

1. When Peking ducklings have been following one of two similar maternal visual replicas, each emitting a different call, and the calls are switched between replicas in the midst of the test, the ducklings either transfer their response to the other replica or completely terminate following the first replica (Gottlieb, 1966). This is a rather complicated result as it stands, and it most certainly warrants further examination using dissimilar replicas, etc., but there is some indication here that in the midst of a conventional imprinting or training situation involving an audiovisual model that (initially, at least) the ducklings are more profoundly influenced by the auditory component than they are by the visual component. This conclusion, if verified across a broader spectrum of conditions, could have important implications for studies of visual imprinting in which the training condition involves an audiovisual model and the testing condition involves only the visual component. In some cases like that, the birds do not choose the "imprinted" object in the later test (e.g., Klopfer, 1971); that is, they do not actually become imprinted to the visual aspect by itself. In an experiment in which the procedure was reversed—where the birds were trained to an audiovisual object and later tested with only the auditory component—the birds did show evidence of auditory imprinting in the later test (Evans & Mattson, 1972). All of these results taken together suggest that in a conventional laboratory imprinting situation the birds may be more profoundly affected by the auditory component of an audiovisual object than by its visual component. This possibility has almost been lost in the traditional overemphasis on the visual side of the problem.

2. So far situations involving only actual duck calls and visual replicas of ducks have been discussed. The early behavioral prepotency of the auditory over the visual system is even more clearly in evidence when artificial or synthetic stimulation is used. For example, chicks select auditory flicker over visual flicker in a simultaneous auditory versus visual choice test, *even when the stimuli have been equated for their attractiveness in singular presentations* (Gottlieb & Simner, 1969). While the previous experiments indicate that the duck calls are more attractive or compelling to the ducklings than the visual replicas of the ducks, the present experiment demonstrates the behavioral prepotency of the auditory system per se over the visual system per se since, in the present case, the evocative nature of the test stimuli has been equated. In the previous experiments, it could be assumed that the calls were merely more evocative than the replicas and nothing more. The present experiment suggests that the prepotency of auditory over visual stimulation in the approach and following response of young ducklings and chicks may represent a fundamental functional difference in the status or role of these two sensory systems at this early stage of postnatal development, at least under given conditions of prior auditory and visual experience.

3. Strong hints of an important functional difference between the auditory and visual modalities have also been present in the work of others on chicks (Fischer, 1966; Smith & Bird, 1963) as well as quail (Porter & Stettner, 1968), even though in those studies no attempt was made to assure the initial equivalence of attractiveness of the stimuli employed. In all of these rather different studies, the young birds demonstrated some kind of positive behavioral preference for the auditory aspect of an audiovisual complex of stimulation. In the face of so much literature dealing with the visual side of imprinting, and the visual side of early experience in general for that matter, it seems significant to conclude that the behavior of the young of the bird species most often used in such research is guided primarily by *auditory* stimulation, at least with regard to the *initial* phase of development of species identification. Let us turn next to a consideration of the ontogenetic factors which could be responsible for such a conclusion.

Neuroembryological and Sensory Stimulative Factors

When we consider the nature of the earliest auditory and visual stimulation encountered by bird embryos, as well as their neuroembryological development, the ontogenetic basis for the early domination of the visual system by the auditory system becomes rather clear. In the chick embryo, for example, there is some neuroanatomical and physiological evidence that the peripheral auditory system may be capable of transmitting impulses at around Day 12–13 of development, which is from 7 to 8 days prior to hatching (evidence reviewed by Gottlieb, 1968). By Day 16 or 17, the chick embryo's auditory system has advanced to the point where a rudimentary kind of auditory conditioning can take place (Gottlieb, 1968). Its visual system, on the other hand, shows the first evidence of function (pupillary reflex) on Day 16–17 (Heaton, 1971), and the first signs of electrical activity in the retina are detected around Day 17–18 (Garcia-Austt & Patetta-Queirolo, 1961; Peters, Vonderahe, & Powers, 1958) or later (Witkovsky, 1963). These and other findings (see Gottlieb, 1968, for a more complete review) indicate the relatively early onset of auditory function and the relatively late onset of visual function in the chick from a neuroembryological perspective. (The earlier functional onset of the auditory system compared to the visual system is not an isolated phenomenon restricted to birds or chicks; the same is true for a variety of mammalian species including humans [literature reviewed by Gottlieb, 1971b].)

The auditory system is not only capable of function earlier than the visual system, but it has a more varied opportunity for such function. Consider, for example, that various sound patterns can penetrate the egg and that both chick and duck embryos can begin stimulating themselves aurally through their own vocalizations several days prior to hatching (Gottlieb & Vandenbergh, 1968). Though light also can penetrate the egg, the embryo's opportunity for a *varied* visual experience inside the egg prior to hatching would seem much less compared to its opportunity for varied auditory stimulation. By way of enhancing the functional differences between the two systems after hatching, it should be noted that most investigators of imprinting keep their hatchlings in visual but not auditory isolation during the time prior to experimentation. To assess the contribution of isolation (deprivation) to the present conclusions, it would be very informative to reverse those rearing conditions and test the birds' auditory and visual preferences using equivalent stimuli as above. According to a point of view expounded in considerable detail elsewhere (Gottlieb, 1971b), the preference for the visual test object should increase under these circumstances, the idea being simply that the visual or auditory responsiveness of the birds is retarded under conditions of isolation and accelerated under conditions of stimulation. In this regard, it is relevant to point out that in every one of the previously cited studies which indicated some sort of positive preference for, or stronger reaction to, the auditory aspect over the visual aspect of an audiovisual object, the birds were maintained in visual but not auditory isolation from the time of hatching to testing (specifically, Fischer, 1966; Gottlieb, 1971a; Gottlieb & Simner, 1969; Porter & Stettner, 1968; Smith & Bird, 1963).

In summary, the relatively late neuroembryological development of visual sensitivity compared to auditory sensitivity, combined with the opportunity for a lesser degree or amount of varied visual stimulation compared to auditory stimulation (both before and after hatching under most circumstances), would seem to conspire to assure the early domination or advancement of auditory perception over visual perception early in avian development. When one harks back to the naturalistic scene (Gottlieb, 1963, 1965), it will perhaps be recalled that the earliest "social" interchange between the incubating parent and the young in the process of hatching is an auditory one (i.e., a vocal interchange), and one would predict that, under these circumstances, the young might be capable of learning the individual peculiarities of their maternal parent's call in advance of learning her particular visual attributes. (Some indirect support for this proposition stems from the observation that guillemots, *Uria* *aalge*, learn to distinguish the individual calls of their parents before they hatch [Tschanz, 1968]. So far we have only weak evidence of the capacity for individual auditory recognition by young ducks [Ramsay, 1951], but it has been demonstrated very clearly in other species [Beer, 1969; Evans, 1970].) Second, under the usual conditions of incubation, hatching, and brooding in nature, and the usual conditions of incubation, hatching, and brooding in the laboratory, it may be that the learning of the visual attributes of the hen (or an artificial visual surrogate) is facilitated or accomplished in the context of exposure to her call. (This proposition receives indirect support from the fact that many experimenters on visual imprinting find it "useful" to have an auditory signal accompany the to-be-imprinted visual object during the initial learning or exposure period.) These predictions are closely tied to the prior neuroembryological history and usual sensory stimulative background of the hatchling. Thus, for example, as far as the latter is concerned, it is important to recall that the usual course of auditory perceptual development in ducklings is itself altered when merely quantitative departures from normally occurring auditory stimulation are introduced during embryonic (Gottlieb, 1971a) or even postnatal phases (Gottlieb, 1966). To be specific, augmentation or premature administration of normally occurring auditory stimulation accelerates auditory perceptual development in embryos as well as hatchlings, while deprivation has the reverse effects on both embryos and hatchlings. A recent study by Reese, Schotté, Bechtold, and Cowley (1972) demonstrates the enormous influence of rearing condition on later choice behavior in young chicks.

Specific and Nonspecific Developmental Background Factors

An imposing and informative task for the future is the disentanglement of specific from nonspecific sensory stimulative background factors in the development of perception. At the moment, it cannot be said for certain whether prior auditory and visual stimulation merely play a rather nonspecific role in the progressive maturation and development of the sensory systems, or whether specific features of the prior stimulation also determine or influence the rather specific perceptual preferences that are evident later on in development (for example, the preference for the maternal call of the species, the preference for certain repetition rates, intensity, etc.).

In order to accomplish such a genuine advance in understanding the development of perception, it is necessary (*a*) to analyze the particular attributes of the stimulating complex to which our experimental subjects are responsive (rate, frequency, modulation, wavelength, internal configuration, intensity, etc.), whether the stimulating complex be auditory or visual and, with this information in hand, (*b*) to try to make a highly specific link with some particular aspect of the bird's prior sensory experience. This challenging task is complicated somewhat by the interesting but as yet unexplored possibility that specific features (e.g., rate) of stimulation in one sensory modality may influence the development of preferences in another modality (e.g., Simner, 1969).

Since the sensory systems become functional sequentially rather than all at once (Gottlieb, 1971b), there is some ontogenetic basis for Schneirla's (1965) hypothesis that certain features of embryonic tactile and proprioceptive stimulation may influence preferences in the later-developing auditory and visual modalities. These kinds of intersensory experiments in birds, which would represent a very advanced stage of inquiry, require a highly analytic understanding of the key attributes of the stimulating complex as a first step; otherwise, there is no basis to establish a link with prior experience in another sensory modality. Of course, the first stage of inquiry would attempt to establish developmental links within rather than between sensory modalities.

A more general way to pose this problem is to ask: Why are certain "training" objects so much more effective than other stimulating complexes? That is, why, *in an ontogenetic sense*, do young chicks or ducklings respond more readily to certain stimulating complexes rather than others upon *initial* presentation? This question—so provocative and so important for our understanding of the development of perception—has been very much neglected in the experimental analysis of imprinting.

SUMMARY AND CONCLUSIONS

The most effective object for eliciting the approach or following response from "naïve" ducklings is one which emits a recording of the call of the hen of the species as she leads the young from the nest in nature, even though the young have not been previously exposed to the hen's call. If an object emits the species-specific maternal call, its visual characteristics can vary widely with respect to size, color, and marking without significantly reducing its attractiveness to the ducklings. If another call is introduced, one with different audiometric characteristics than the species' maternal call, all visual objects tested become significantly less attractive.

As the results of auditory choice tests show, the birds come into the imprinting situation already "tuned" to the distinctive audiometric characteristics of the maternal assembly call of their own species; they selectively respond to it in preference to other calls which have different audiometric characteristics. This is not the case, however, with a purely visual replica of a hen of the species, nor even an appropriately vocalizing replica of the hen—the ducklings are not selectively responsive to a visual replica of a hen of their own species, even when it is emitting the species' maternal call. Thus, species identification in ducklings is initially founded on auditory perception and, in nature, the learning of the visual characteristics of the hen takes place in the context of the duckling's response to the call of the hen.

From a neuroembryological perspective, the auditory system of birds (and other animals) develops in advance of their visual system, which leads to a behavioral domination of the visual system by the auditory system, at least early in development. While neural maturation processes are involved—that is, a different onset or rate of maturation of the two sensory systems—the sensory stimulative environment may also contribute to the early disparity between auditory and visual function. In most imprinting studies, for example, the hatchlings are maintained in visual but not auditory isolation prior to their entrance into the imprinting situation.

The initial, so-called naïve perceptual preferences of animals have long been overlooked as a fruitful starting point for developmental study; this is nowhere more evident than in the literature on avian imprinting where there is, in fact, the least justification for such an oversight. Namely, there is now considerable evidence to show that the auditory and visual systems of chicks, ducklings, and other avian species are functional prior to hatching, and stimulation during the embryonic period is known to affect the behavior of the embryo as well as that of the hatchling, so attribution of perceptual naïveté to young birds which have been communally incubated and brooded in noisy and illuminated surroundings hardly seems reasonable. In fact, the failure to closely control or monitor the sheer amount and kind of preexperimental auditory or visual stimulation, both before and after hatching, ignores the demonstrated effects of these variables on behavior in imprinting-type situations (e.g., Bateson, 1964; Dimond, 1968; Gottlieb, 1966, & 1971a; Gottlieb & Klopfer, 1962; Grier, Counter, & Shearer, 1967; Haywood, 1965; Impekoven, 1973. A fully developmental understanding of behavior needs to take into account the sensory environment and perceptual capabilities of the embryo as well as the hatchling.

REFERENCES

BATESON, P. P. G. Effect of similarity between rearing and testing conditions on chicks' following and avoidance responses. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 100–103.

BEER, C. G. Laughing gull chicks: Recognition of their parents' voices. *Science*, 1969, **166**, 1030–1032.

BJÄRVALL, A. The hatching and nest-exodus behaviour of Mallard. *Wildfowl*, 1968, **19**, 70–80.

BOYD, H., & FABRICIUS, E. Observations on the incidence of following of visual and auditory stimuli in naive Mallard ducklings (*Anas platyrhynchos*). *Behaviour*, 1965, **25**, 1–15.

COLLIAS, N. E., & COLLIAS, E. C. Some mechanisms of family integration in ducks. *Auk*, 1956, **73**, 378–400.

DIMOND, S. J. Effects of photic stimulation before hatching on the development of fear in chicks. *Journal of Comparative and Physiological Psychology*, 1968, **65**, 320–324.

EVANS, R. M. Parental recognition and the "mew call" in black-billed gulls (*Larus bulleri*). *Auk*, 1970, **87**, 503–513.

Evans, R. M., & Mattson, M. E. Development of selective responses to individual maternal vocalizations in young *Gallus gallus*. *Canadian Journal of Zoology*, 1972, **50**, 777–780.

Fabricius, E. The behavior of young Anatidae. *Acta Zoologica Fennica*, 1951, **68**, 1–178.

Fischer, G. J. Auditory stimuli in imprinting. *Journal of Comparative and Physiological Psychology*, 1966, **61**, 271–273.

Frith, H. J. *The Mallee-fowl*. London: Angus & Robertson, 1962.

Garcia-Austt, E., & Patetta-Queirolo, M. A. Electroretinogram of the chick embryo. I. Onset and development. *Acta Neurologica Latinoamericana*, 1961, **7**, 179–189.

Gottlieb, G. A naturalistic study of imprinting in Wood ducklings (*Aix sponsa*). *Journal of Comparative and Physiological Psychology*, 1963, **56**, 86–91.

Gottlieb, G. Components of recognition in ducklings. *Natural History*, 1965, **74**, 12–19.

Gottlieb, G. Species identification by avian neonates: Contributory effect of perinatal auditory stimulation. *Animal Behaviour*, 1966, **14**, 282–290.

Gottlieb, G. Prenatal behavior of birds. *Quarterly Review of Biology*, 1968, **43**, 148–174.

Gottlieb, G. *Development of species identification in birds: An inquiry into the prenatal determinants of perception*. Chicago: University of Chicago Press, 1971. (a)

Gottlieb, G. Ontogenesis of sensory function in birds and mammals. In E. Tobach, L. R. Aronson, & E. Shaw (Eds.), *The biopsychology of development*. New York: Academic Press, 1971. (b)

Gottlieb, G., & Klopfer, P. H. The relation of developmental age to auditory and visual imprinting. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 821–826.

Gottlieb, G., & Simner, M. L. Auditory versus visual flicker in directing the approach response of domestic chicks. *Journal of Comparative and Physiological Psychology*, 1969, **67**, 58–63.

Gottlieb, G., & Vandenbergh, J. G. Ontogeny of vocalization in duck and chick embryos. *Journal of Experimental Zoology*, 1968, **168**, 307–325.

Grier, J. B., Counter, S. A., & Shearer, W. M. Prenatal auditory imprinting in chickens. *Science*, 1967, **155**, 1692–1693.

Haywood, H. C. Discrimination and following behavior in chicks as a function of early environmental complexity. *Perceptual and Motor Skills*, 1965, **21**, 299–304.

Heaton, M. B. Ontogeny of vision in the Peking duck (*Anas platyrhynchos*): The pupillary light reflex as a means for investigating visual onset and development in avian embryos. *Developmental Psychobiology*, 1971, **4**, 313–332.

Impekoven, M. Response-contingent prenatal experience of maternal calls in the Peking duck (*Anas platyrhynchos*). *Animal Behaviour*, 1973, **21**, 164–168.

Klopfer, P. H. Imprinting: Determining its perceptual basis in ducklings. *Journal of Comparative and Physiological Psychology*, 1971, **75**, 378–385.

Lorenz, K. Der Kumpan in der Umwelt des Vogels. *Journal für Ornithologie*, 1935, **83**, 137–213, 289–413.

Lorenz, K. The companion in the bird's world. *Auk*, 1937, **54**, 245–273.

Peters, J. J., Vonderahe, A. R., & Powers, T. H. Electrical studies of functional development of the eye and optic lobes in the chick embryo. *Journal of Experimental Zoology*, 1958, **139**, 459–468.

Porter, R. H., & Stettner, L. J. Visual and auditory influences on following responses of Bobwhite quail (*Colinus virginianus*). *Journal of Comparative and Physiological Psychology*, 1968, **66**, 808–811.

Ramsay, A. O. Familial recognition in domestic birds. *Auk*, 1951, **68**, 1–16.

Reese, E. P., Schotté, C. S., Bechtold, R. E., & Cowley, V. L. Initial preference of chicks from five rearing conditions for a hen or a rotating light. *Journal of Comparative and Physiological Psychology*, 1972, **81**, 76–83.

Schneirla, T. C. Aspects of stimulation and organization in approach/withdrawal processes underlying vertebrate behavioral development. In D. S. Lehrman, R. A. Hinde, & E. Shaw (Eds.), *Advances in the study of behavior*. Vol. 1. New York: Academic Press, 1965.

Simner, M. L. The cardiac self-stimulation hypothesis and nonnutritive sucking in human infants. *Developmental Psychology*, 1969, **1**, 569–575.

Smith, F. V., & Bird, M. W. The relative attraction for the domestic chick of combinations of stimuli in different sensory modalities. *Animal Behaviour*, 1963, **11**, 300–305.

Tschanz, B. Trottellummen, Die Entstehungen der persönlichen Beziehungen zwischen Jungvogel und Eltern. *Zeitschrift für Tierpsychologie*, 1968, 1–103. (Suppl. 4)

Witkovsky, P. An ontogenetic study of retinal function in the chick. *Vision Research*, 1963, **3**, 341–355.

(Received March 29, 1972)

Psychological Bulletin
1973, Vol. 79, No. 6, 373–375

NONNECESSARY CONDITIONS OR USEFUL PROCEDURES IN DESENSITIZATION:

A REPLY TO WILKINS

WESLEY G. MORGAN[1]

University of Tennessee

Muscle relaxation, a graded hierarchy of fear-relevant scenes and imagination of these scenes concommitant to muscle relaxation are three procedures defining Wolpe's Systematic Desensitization which Wilkins claims are unnecessary. These claims follow from a somewhat prejudiced review of largely analogue studies and may be misleading to the practicing behavior therapist. Explanations of the success of desensitization in terms of cognitive and social variables seem to have no more support or utility than Wolpe's original theoretical formulations. It would seem premature to discard the defining procedures of desensitization as unnecessary when they are quite useful and perhaps necessary for success in treating highly fearful patients.

Behavior therapists may be justly troubled on research, theoretical, and practical grounds by the conclusions and suggestions offered by Wilkins (1971). Wilkins states that the effectiveness of the systematic desensitization package (which he does not seem to dispute) seems

> due to social variables involved in the patient-therapist relationship and cognitive variables involving expectancy of therapeutic gain, information feedback of success, training in the control of attention, and vicarious learning of the contingencies of behavior through instructed imagination [p. 311].

This view has the unfortunate effect of minimizing extremely important procedures and operations while substituting for them concepts which are poorly defined operationally and for which comparatively little support exists.

NECESSITY OF MUSCLE RELAXATION

Wilkins (1971) is correct in asserting that deep muscle relaxation is not a *necessary* condition for "desensitization" to take place if one agrees to mean by "desensitization" a diminution in fear behavior rather than the defining operations of Wolpe's procedure. However, it is important to point out the several senses in which muscle relaxation is not *necessary*. (*a*) It is not *necessary* in the sense that other anxiety-inhibiting responses or conditions may be used as well. In fact, any condition or response which is incompatible with the experience of anxiety would be acceptable for use in the counterconditioning procedure as Wolpe has outlined it. For example, Wolpe and Lazarus (1966) report the use of emotive imagery, feeding responses, conditioned motor responses, and sexual responses as anxiety-inhibiting operations, and there is no reason to believe that this exhausts the possibilities. (*b*) It is not *necessary* in the sense that one could become desensitized to a stimulus through simple extinction (omitting the anxiety-inhibiting operations) as well as through counterconditioning procedures (Malleson, 1959; Stampfl & Levis, 1967; Wolpin & Raines, 1966). However, the practicality and comparative utility of these other techniques for severely phobic patients still needs support. And (*c*), not independent of *b* above, relaxation may not be a *necessary* condition for desensitization to take place with subjects demonstrating a weak or mild anxiety as opposed to a severe anxiety.

In contrast to papers cited by Wilkins (1971), other studies with human subjects have shown facilitative effects of relaxation in the reduction of fears of nonclinical proportions. (Davison [1965],[2] 1968; Rachman, 1965). Cooke (1966) has shown that subjects with higher Bendig Emotionality Scores improved more when their treatment employed relaxation than when it did not, and Schubot[3] was also able to demonstrate that relaxation was very important in modifying the fears of highly fearful subjects while it did not facilitate the reduction of fear in moderately fearful subjects. Thus the importance of relaxation may be positively correlated with the intensity of the original fear and would be of relatively greater importance to the clinical practitioner than to the analogue researcher or theoretician.

---

[1] Requests for reprints should be sent to Wesley G. Morgan, Psychological Clinic, Department of Psychology, University of Tennessee, Knoxville, Tennessee 37916.

NECESSITY OF A GRADED HIERARCHY OF FEAR-RELEVANT SCENES

Wilkins (1971) has claimed that individually tailored hierarchies of feared scenes are unnecessary for successful desensitization. The experimental studies cited in support of this claim have been analogue studies, and one must consider how relevant these studies have been before accepting this conclusion.

All the analogue studies have used spatial–temporal hierarchies (Paul, 1969) while those problems encountered in applied clinical situations are much more likely to involve thematic hierarchies. (Hain, Butcher, & Stevenson, 1966; Marquis & Morgan, 1968; Wolpe, 1961). With these more complex kinds of hierarchies numerous graded scenes are more likely to be required. There is also reasonable doubt that the participating subjects actually were as severely fearful as one would likely obtain in a clinical case. As Bernstein and Paul (1971) have noted, "To the extent that strong anxiety responses are *not* elicited, the need for lower hierarchy items at all, let alone their order, duration, manner of presentation, and so forth, becomes theoretically irrelevant [p. 233]."

IMAGINATION OF FEAR-RELEVANT SCENES

Wilkins (1971) states, "Indeed, the only necessary condition appears to be imagination of fear-relevant scenes not necessarily arranged into a hierarchy and not necessarily concommitant to muscle relaxation [p. 311]." If one were to seriously take this position, it would be encumbant on him to explain why patients with nightmares and obsessions concerning feared situations are not always "spontaneously" cured. Surely there must be other factors of influence. Wilkins suggests three other factors which he claims operate in the desensitization situation: (*a*) expectancy of therapeutic gain, (*b*) the therapist as a reinforcer of nonfear behavior, and (*c*) the role of information feedback, all of which he claims are highly influential in affecting outcome.

EXPECTANCY OF THERAPEUTIC GAIN

Wilkins (1971) argued that the expectancy of therapeutic gain contributed "a large part" to the actual gain obtained in systematic desensitization. Since then, in a more carefully prepared critique of this area, Wilkins (1973) concludes, "that the construct 'expectancy of therapeutic gain' emerged prematurely and without the empirical support necessary to establish its validity." It is evident from studies cited in this more recent critique and others (Borkovec, 1972; Davison (see Footnote 1); Lang, Lazovik, & Reynolds, 1965; McGlynn, 1971) that while one cannot state that expectancy plays no role in the outcome of desensitization (one cannot affirm the null hypothesis), one can conclude that it has proved extremely difficult to demonstrate in the context of systematic desensitization, and that it most probably plays a minimal role at best in the achievement of obtained results.

THERAPIST AS A REINFORCER OF NONFEAR BEHAVIOR AND THE ROLE OF INFORMATION FEEDBACK

Most would not wish to quarrel with the effectiveness of contingent reinforcement, whether in the form of social reinforcement or information feedback, for increasing the probability of adaptive, appropriate behavior. These operant techniques can, and probably do, add significantly to the effectiveness and efficiency of the desensitization package. But, in the absence of direct demonstrations, to claim that these are "critical variables" for successful therapeutic outcome in desensitization seems to go considerably beyond the data available, and to suggest that the major operations of Wolpe's systematic desensitization procedure are not useful in the actual treatment of severely fearful patients would be quite premature.

---

[2] G. C. Davison. The influence of systematic desensitization, relaxation, and graded exposure to imaginal aversive stimuli on the modification of phobic behavior. Unpublished doctoral dissertation, Stanford University, 1965.

[3] E. D. Schubot. The influence of hypnotic and muscular relaxation in systematic desensitization of phobias. Unpublished doctoral dissertation, Stanford University, 1966.

REFERENCES

BERNSTEIN, D. A., & PAUL, G. L. Some comments on therapy analogue research with small animal "phobias." *Journal of Behavior Therapy and Experimental Psychiatry*, 1971, **2**, 225–237.

BORKOVEC, T. D. Effects of expectancy on the outcome of systematic desensitization and implosive treatments for analogue anxiety. *Behavior Therapy*, 1972, **3**, 29–40.

COOKE, G. The efficacy of two desensitization procedures: An analogue study. *Behaviour Research and Therapy*, 1966, **4**, 17–24.

DAVISON, G. C. Systematic desensitization as a counterconditioning process. *Journal of Abnormal Psychology*, 1968, **73**, 91–99.

HAIN, J., BUTCHER, R., & STEVENSON, I. Systematic desensitization therapy; an analysis of results in twenty-seven patients. *British Journal of Psychiatry*, 1966, **112**, 295–307.

LANG, P. J., LAZOVIK, A. D., & REYNOLDS, D. J. Desensitization, suggestibility, and pseudotherapy. *Journal of Abnormal Psychology*, 1965, **70**, 395–402.

MALLESON, N. Panic and phobia. *Lancet*, 1959, **1**, 225–227.

MARQUIS, J. N., & MORGAN, W. G. *A guidebook for systematic desensitization*. Palo Alto, Calif.: Veterans' Workshop, 1968.

MCGLYNN, F. D. Experimental desensitization following three types of instructions. *Behaviour Research and Therapy*, 1971, **9**, 367–369.

PAUL, G. L. Outcome of systematic desensitization I: Background procedures, and uncontrolled reports of individual treatment. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status*. New York: McGraw-Hill, 1969.

RACHMAN, S. Studies in desensitization—I: the separate effects of relaxation and desensitization. *Behaviour Research and Therapy*, 1965, **3**, 245–252.

STAMPFL, T., & LEVIS, D. Essentials of implosive therapy. *Journal of Abnormal Psychology*, 1967, **72**, 496–503.

WILKINS, W. Desensitization: Social and cognitive factors underlying the effectiveness of Wolpe's procedure. *Psychological Bulletin*, 1971, **76**, 311–317.

WILKINS, W. Expectancy of therapeutic gain: An empirical and conceptual critique. *Journal of Consulting and Clinical Psychology*, 1973, **40**, 69–77.

WOLPE, J. The systematic desensitization treatment of neuroses. *Journal of Nervous and Mental Disease*, 1961, **132**, 189–203.

WOLPE, J., & LAZARUS, A. *Behavior therapy techniques*. New York: Pergamon, 1966.

WOLPIN, M., & RAINES, J. Visual imagery, expected roles and extinction as possible factors in reducing fear and avoidance behavior. *Behaviour Research and Therapy*, 1966, **4**, 25–37.

(Received May 1, 1972)

Psychological Bulletin
1973, Vol. 79, No. 6, 376–377

DESENSITIZATION:

A REJOINDER TO MORGAN

WALLACE WILKINS [1]

University of Miami

Specific points raised by Morgan are debatable on empirical and logical grounds; however, the distinction between necessary and useful operations provides a meaningful framework for the advancement of desensitization theory, practice, and research.

The major controversy involving systematic desensitization does not concern its effectiveness (Paul, 1969a, 1969b), rather the mechanisms responsible for fear reduction. By distinguishing between necessary and useful elements of Wolpe's (1958) procedure, Morgan (1973) has facilitated the search for basic mechanisms.

NECESSARY OPERATIONS

To meet the defining operation of pairing imagination with relaxation, three types of instructions are used: inducing muscle relaxation, directing attention toward scenes, and shifting attention from scenes. Since, as Morgan (1973) agrees, muscle relaxation is not necessary for fear reduction, instructions directing attention toward and from fear-related scenes remain the only necessary operations. This *necessity* of scene imagination does not logically mean, as Morgan (1973) implies is Wilkins' (1971) position, that imagination is also *sufficient* for fear reduction.

USEFUL OPERATIONS

Morgan (1973) states that muscle relaxation is not necessary for fear reduction when simple extinction is used. However, in the articles cited as using "simple extinction," not only are anxiety-inhibiting operations omitted but anxiety-maximizing operations are intentionally included. As such, Morgan's (1973) simple extinction argument becomes an uninformative tautology: Procedures which inhibit anxiety are not necessary if procedures which elicit anxiety are used.

If muscle relaxation is necessary in treating severe anxiety (as suggested by Morgan 1973), the implosion procedure should not be effective with severe anxiety. However, since reports by Hogan (1966, 1968) attest to its effectiveness in treating clinical cases, instructed muscle relaxation does not appear necessary even in treating severe anxiety.

While neither feeding, motor, sexual, nor relaxation responses (Morgan, 1973) are necessary for fear reduction, each is useful and each may function as a tactical maneuver (Davison & Wilson, 1972) by providing a direction for attention when attention is shifted from imaginal scenes.

DIRECTIONS FOR RESEARCH

The possibility emerges that any procedure which contains instructions to shift attention toward and from symptom-related scenes will facilitate symptom reduction. However, since instructed attention shifts may not be sufficient for improvement, a delineation of the additional necessary and useful operations surrounding instructed attention shifts is in order.

Morgan (1973) calls for the validation of results of analog research on clinical populations and Paul (1969b) calls for the exploration of therapist variables in desensitization. It appears important to identify under which conditions and for which populations are which procedures most useful. Studies by Evans (1968), Farmer and Wright (1971), Grim (1971), and Lang, Melamed, and Hart (1970) exemplify this research direction.

---

[1] The author wishes to thank Wesley Morgan for the constructive comments offered during the preparation of this paper.

Requests for reprints should be sent to Wallace Wilkins, Clinical Psychology Training Program, Department of Psychology, University of Miami, Coral Gables, Florida 33124.

REFERENCES

DAVISON, G. C., & WILSON, G. T. Critique of "Desensitization: Social and cognitive factors underlying the effectiveness of Wolpe's procedure." *Psychological Bulletin,* 1972, **78**, 28–31.

EVANS, D. R. Masturbatory fantasy and sexual deviation. *Behaviour Research and Therapy,* 1968, **6**, 17–19.

FARMER, R. G., & WRIGHT, J. M. C. Muscular reactivity and systematic desensitization. *Behavior Therapy,* 1971, **2**, 1–10.

GRIM, P. F. Anxiety change produced by self-induced muscle tension and by relaxation with respiration feedback. *Behavior Therapy,* 1971, **2**, 11–17.

HOGAN, R. A. Implosive therapy in the short term treatment of psychotics. *Psychotherapy: Theory, Research and Practice,* 1966, **3**, 25–32.

HOGAN, R. A. The implosive technique. *Behaviour Research and Therapy,* 1968, **6**, 423–431.

LANG, P. J., MELAMED, B. G., & HART, J. A psychophysiological analysis of fear modification using an automated desensitization procedure. *Journal of Abnormal Psychology,* 1970, **76**, 220–234.

MORGAN, W. G. Non-necessary conditions or useful procedures in desensitization: A reply to Wilkins. *Psychological Bulletin,* 1973, **79**, 373–375.

PAUL, G. L. Outcome of systematic desensitization. I: Background, procedures, and uncontrolled reports of individual treatment. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status.* New York: McGraw-Hill, 1969. (a)

PAUL, G. L. Outcome of systematic desensitization. II: Controlled investigations of individual treatment, technique variations, and current status. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status.* New York: McGraw-Hill, 1969. (b)

WILKINS, W. Desensitization: Social and cognitive factors underlying the effectiveness of Wolpe's procedure. *Psychological Bulletin,* 1971, **76**, 311–317.

WOLPE, J. *Psychotherapy by reciprocal inhibition.* Stanford, California: Stanford University Press, 1958.

(Received November 7, 1972)

Psychological Bulletin
1973, Vol. 79, No. 6, 378–379

GAMBLE'S REVIEW OF THE HOLTZMAN INKBLOT TECHNIQUE:

CORRECTIONS AND CLARIFICATIONS

JON D. SWARTZ[1]

University of Texas at Austin

Certain statements by Gamble regarding results obtained with the Holtzman Inkblot Technique (HIT), and the relationship of the HIT to the Rorschach technique, are presented and corrected. Specifically, previous results obtained for the HIT Pathognomic Verbalization (V) score are clarified; the distinctions between Rorschach and HIT Human, Movement, and Color scores are presented. Some recent extensions of HIT research are discussed. Finally, two other minor errors are pointed out and rectified.

Gamble (1972) has provided a generally good, basic review of the Holtzman Inkblot Technique (HIT). Given the unique status of the *Psychological Bulletin* in psychology, his review probably will influence those working with the HIT for many years to come; unfortunately, several of his statements either are in error or in need of some clarification.

In an evaluative statement concerning results obtained for the HIT Pathognomic Verbalization (V) score in two studies of developmental level carried out by Thorpe and Swartz (1965, 1966), Gamble reported that "V first declined and then rose again" in the first study and that "at variance with the earlier data, V was found to decrease with age [p. 179]" in the follow-up study. In point of fact, V scores declined from a mean of 11.4 at age 5 years to 5.6 at ages 8–9, to a low of 2.6 at ages 12–14, and then rose to mean scores of 6.6 and 6.1 at ages 16 and 18–19 years, respectively, in the initial study; in the follow-up investigation mean V scores declined from a mean of 7.4 at age 6.7 years to 5.1 at age 9.7 to a low of 2.3 at age 12.7 years, the oldest criterion group in the follow-up study. A low score of 2.6 at age 12.7 years in the initial study is hardly "at variance" with a low score of 2.3 at age 12.7 years in the follow-up study. In addition, in a cross-cultural validation of the results of these first two studies of HIT perceptual development scores, it was found that 100 Mexican school children, also at age 12.7 years, had a mean V score of 2.5 (Swartz, Lara Tapia, & Thorpe, 1967). In all HIT studies of normal subjects carried out to date, V scores decline dramatically with age until around 12–13 years, at which time they rise again. The qualitative nature of the *categories* of Pathognomic Verbalization are strikingly different in V scores produced after puberty, however. With increasing age it has been found that there is a significant increase in the number of children giving fabulized responses and significant decreases in the number giving autistic logic, contaminated, and absurd responses.[2]

In concluding his discussion of developmental studies using the HIT, Gamble (1972) suggests that a next logical step

> would involve testing subjects in middle and old age to determine whether the HIT is sensitive to the reverse trends of decreasing differentiation and hierarchic integration predicted by organismic developmental theory [p. 180].

The study of middle-aged subjects already has been carried out (Witzke, Swartz, & Drew, 1971), and the study of subjects in old age currently is in progress.[3] In addition, a recent investigation of the relationships between mental age and the several HIT scores that have proved to be reliable and meaningful indexes of perceptual development also provided strong support for these HIT variables as indexes of perceptual development and confirmed the nature and the direction of this development (Swartz, Cleland, Drew, & Witzke, 1971).

---

[1] Requests for reprints should be sent to Jon D. Swartz, 310A W. C. Hogg Building, the University of Texas at Austin, Austin, Texas 78712. The author wishes to thank Wayne H. Holtzman for his helpful comments on an earlier draft of the paper.

[2] J. D. Swartz. Pathognomic verbalizations in normals, psychotics, and mental retardates. (Doctoral dissertation, University of Texas at Austin) Ann Arbor, Mich.: University Microfilms, 1969. No. 70-10, 872.

[3] Keith F. Bell is a coinvestigator in this study.

The clarifications that need to be made regarding the relationship of the HIT with the Rorschach involve the HIT scores of Movement, Human, and Color:

1. Gamble (1972) states that "Rorschach Human Movement (*M*) is [HIT] Movement scores 2 or higher [p. 175]." In fact, Rorschach Human Movement could be HIT Movement 1, 2, 3, or 4, depending upon the response. In the HIT scoring system, a Movement score of 1 is reserved for those responses with "static potential for movement" (sitting, looking, etc.) (Holtzman, Thorpe, Swartz, & Herron, 1961, p. 51). According to at least one Rorschach authority, such potential would be scored Human Movement since "human movement is used to score concepts involving any kind of humanlike action, posture, or expression [Klopfer & Davidson, 1962, p. 70]."

2. Related to Point 1, Gamble then states that "Rorschach Entire Human (*H*) is [HIT] Human 1 or 2 [p. 175]." This statement also is in error. Rorschach Entire Human would most often be HIT Human 2 while Rorschach Parts of Human Figures (*Hd*) would most often be HIT Human 1. Most other Rorschach content scores related to human content [(*H*), *AH*] also would be scored HIT Human 1.

3. When Gamble writes further that "Rorschach *FC*, *CF*, or *C* would correspond to the number of [HIT] *chromatic* (italics added) cards coded 1, 2, or 3, respectively, for Color [p. 175]," he is also at least partly in error. In the HIT scoring system no distinction is made between chromatic and achromatic cards in scoring Color. It is quite possible, therefore, for HIT Color scores to be given to completely achromatic cards. It is true that Rorschach *FC*, *CF*, or *C* probably would correspond very closely to the number of responses to chromatic *areas* of HIT cards coded 1, 2, or 3, respectively, however.

4. Finally, two minor errors of a typographical nature might lead to some confusion if not corrected: (*a*) In listing the nine qualitative Pathognomic Verbalization (V) categories in the HIT scoring system, Gamble (1972) refers to "deteriorated responses [p. 174]." Most V scores, regardless of qualitative category, can be thought of as "deteriorated." Gamble intended to write "deteriorated *color*" responses. (*b*) When Gamble writes "Rorschach Whole (*W*), *Determinants* (italics added) (*D*), or Common Detail (*d*) would be [HIT] Location 0, 1, or 2, respectively [p. 175]," he do doubt meant to write "Large Usual Detail" instead of "Determinants."

REFERENCES

GAMBLE, K. R. The Holtzman Inkblot Technique: A review. *Psychological Bulletin,* 1972, 77, 172–194.

HOLTZMAN, W. H., THORPE, J. S., SWARTZ, J. D., & HERRON, E. W. *Inkblot perception and personality.* Austin: University of Texas Press, 1961.

KLOPFER, B., & DAVIDSON, H. H. *The Rorschach technique.* New York: Harcourt, Brace & World, 1962.

SWARTZ, J. D., CLELAND, C. C., DREW, C. J., & WITZKE, D. B. The Holtzman Inkblot Technique as a measure of perceptual development in mental retardation. *Journal of Personality Assessment,* 1971, 35, 320–325.

SWARTZ, J. D., LARA TAPIA, L., & THORPE, J. S. Perceptual development of Mexican school children as measured by responses to the Holtzman Inkblot Technique. *Revista Interamericana de Psicologia,* 1967, 1, 289–295.

THORPE, J. S., & SWARTZ, J. D. Level of perceptual development as reflected in responses to the Holtzman Inkblot Technique. *Journal of Projective Techniques and Personality Assessment,* 1965, 29, 380–386.

THORPE, J. S., & SWARTZ, J. D. Perceptual organization: A developmental analysis by means of the Holtzman Inkblot Technique. *Journal of Projective Techniques and Personality Assessment,* 1966, 30, 447–451.

WITZKE, D. B., SWARTZ, J. D., & DREW, C. J. Level of perceptual development of normal adults as measured by the Holtzman Inkblot Technique. *Proceedings of the 79th Annual Convention of the American Psychological Association,* 1971, 6, 609–610.

(Received May 18, 1972)

Psychological Bulletin
1973, Vol. 79, No. 6, 380–385

PROBLEMS IN THE MEASUREMENT OF COGNITIVE DEFICIT[1]

LOREN J. CHAPMAN[2] AND JEAN P. CHAPMAN

University of Wisconsin

Several dozen studies of differential cognitive deficit appear each year. Artifacts cloud the findings in a majority of these studies because of differences in discriminating power of tasks coupled with generalized deficit of the patients. With two tests of differing reliability, the test with the higher reliability will yield greater performance deficit for the less able subjects. We argue for the advisability, in studies of differential cognitive deficit, of matching tasks on reliability, shape of the distribution of scores, and mean, variance, and shape of the distribution of item difficulty, using normal subjects alone as a standardization group.

Most hypotheses about cognitive deficit are tested in designs in which two or more error scores are compared. We show that in such studies artifactual findings are produced by differences in test characteristics which affect discriminating power. We illustrate these principles by studies of schizophrenic thought disorder, although the principles apply equally well to hypotheses about all other cognitive pathologies such as those of brain damage, drug states, and mental deficiency.

Most investigations of cognitive deficit are concerned with specific deficits. A single score per subject is not sufficient to measure a specific deficit because subjects with cognitive pathology typically show generalized cognitive deficit, that is, low scores on almost any measure of intellectual functioning. Therefore, meaningful statements about specific deficit must be in terms of differential deficit, that is, a greater loss in one ability than in one or more other abilities.

DISCRIMINATING POWER AND MEASURES OF DIFFERENTIAL DEFICIT

The degree of differential deficit in ability is inferred, of course, from a comparison of the extent to which two measures discriminate pathological performance from normal performance. However, the extent of the inferiority of score of pathological subjects depends not only on their deficit in ability but also on the discriminating power of the test. "Discriminating power" refers here to the extent to which the score differentiates the more able from the less able subjects and, hence, differentiates two groups that differ in the ability measured by the test. Tests have identical discriminating power if they have identical distributions of error-free true scores. If deficient subjects are as inferior to normal subjects on one ability as on another, but the test that is used to measure one of the abilities is more discriminating than the test of the other, a greater performance deficit will be found on the more discriminating measure.

For tests of the same format and mode of scoring (e.g., free response items with dichotomous scoring), the discriminating power of a test is a function of mean item difficulty, dispersion of item difficulty, mean item covariance, and number of items. We focus first on item difficulty. In many published studies of cognitive pathology the pairs of tasks used differ sharply on item difficulty in ways which

---

[1] Preparation of this article was supported by a research grant (MH-18354) and by a Research Scientist Award (K05-MH-05198) to the first author; both grants are from the National Institute of Mental Health.

The authors are indebted to Anne Cleary, Jum Nunnally, and Lyle Jones for criticisms of an earlier draft of this manuscript, and to Randall Daut and Sonja Farley for assistance in gathering the data. They are also indebted to Janet Rafferty for the data on school children; to Joseph Cocke, Richard Cameron, and Thomas Pritchett of San Antonio State Hospital for permission to test schizophrenic subjects; and to Elmer Cady, Raymond Anderson, and Thomas Biever of Wisconsin State Prison for permission to test prison inmates.

[2] Requests for reprints should be sent to Loren J. Chapman, Psychology Department, University of Wisconsin, Madison, Wisconsin 53706.

could easily have produced the results of the study artifactually.

Other things being equal, items closer to the 50% level of difficulty for all subjects tested have the highest discriminating power on a dichotomously scored free response task composed of items on which guessing is not likely to produce the correct answer (Lord, 1952b). On multiple-choice tasks, the highest discriminating power is expected for accuracy levels slightly higher than midway between chance and 100% accuracy (Lord, 1952a).

The effect of item difficulty on discrimination of groups of differing ability can be illustrated by comparing two age groups of children on vocabulary items that differ in difficulty. Figure 1 shows the percentage accuracy of a group of 34 third-grade children and a group of 39 fifth-grade children on the vocabulary items of the Stanford-Binet Intelligence Scale. The abscissa of Figure 1 presents six sequential groups of vocabulary items, each representing a different level of difficulty (as measured by percentage accuracy for third and fifth grades combined). Each point on the abscissa represents several vocabulary items of a given level of difficulty. Against this abscissa are plotted curves for the percentage of third-grade children and the percentage of fifth-grade children passing each set of items. (If the two grades were combined, the resultant curve would represent a variable plotted against itself, except for the slightly different scale conventions of the abscissa and ordinate.) The curves for the two grades tend to join at the extreme ends of the dimension of difficulty, and they attain greatest separation for items at about the 50% level of difficulty for the two groups combined. This relationship between difficulty and discriminating power may be generalized to any two groups that differ in accuracy, including pathological and normal subjects.

To illustrate how differences in difficulty can be mistaken for differential deficit, let us examine a study by Truscott (1970). Truscott compared schizophrenic and normal subjects on four tasks, including recall of random word strings and recall of normal sentences. The groups differed less on recall of random word strings than on recall of normal sentences. Truscott concluded that schizophrenics show a lack of facilitation of recall by normal syntactic constraints.

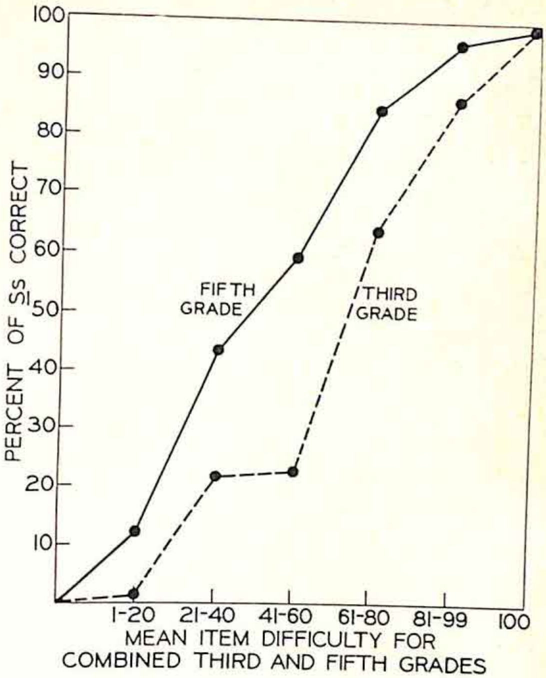


FIG. 1. Accuracy of third- and fifth-grade children on Stanford-Binet vocabulary items of gradated difficulty.

However, her results may instead have been at least in part an artifact of the levels of difficulty (and, hence, discriminating power) of her two tasks. Truscott's tasks were very difficult (corresponding to the left-hand side of Figure 1). Recall of random word strings was at about 20% accuracy for the two groups combined and recall of sentences was much closer to 50% accuracy. If Truscott had used shorter word sequences for both tasks, she would presumably have sampled items at the easy end of the difficulty dimension (corresponding to the right-hand side of Figure 1). Because random word strings are often more difficult than sentences, they would then have been closer than sentences to the middle range of difficulty. Hence, Truscott would probably have found the opposite result, that is, a greater deficit in recalling random word strings than in recalling sentences.

In many studies one cannot judge the direction of the probable artifact because it is not clear which task is more discriminating for normal subjects. If the items of both tasks are very easy, most normal subjects make almost no errors on either task, but a less able group that makes more errors might find one task more difficult than the other. Still other studies use tasks for which adequacy of performance is defined by criteria other than number of errors, such as by speed of response, as in reaction time tasks, or trials to criterion, as in verbal learning. If the tasks are not matched on these criteria of difficulty, the direction of the probable artifact is ambiguous because the most discriminating level of performance is unknown. One may nevertheless be sure that differences in difficulty produce differential discrimination of groups.

How to Match Tests on Discriminating Power

In conventional psychometric theory, reliability is the index of discriminating power in that the square root of the reliability is the correlation of obtained scores with error-free true scores. Yet equivalent reliability is not sufficient by itself for matching tests in studies of differential deficit because of differences in general ability level between the standardization sample and the experimental subjects. The rationale follows.

The proper standardization sample consists of normal subjects only. Tests differ in discriminating power to the extent that they distinguish the more able from the less able subjects among all subjects tested, that is, the normal and pathological subjects treated as a single group. Yet, one would not wish to match tasks for the two groups combined because reliability of the tests is affected by any true differential deficit due to pathology. The task with the greater schizophrenic deficit will have the larger variance for the two groups combined, and hence, the higher reliability. Matching on reliability for the two groups combined would eliminate some or all of this evidence of differential deficit in ability.

Tasks should instead be matched using normal subjects of a wide range of ability. If the tasks are developed before the experiment proper, the groups in the experiment will probably have a different range of ability than the standardization sample. As a result, mean item difficulty of the tests will be different than in the standardization sample, so that tests that had been matched only on reliability and not on item difficulty would probably become unmatched on reliability for the new groups. However, if tests are also initially matched on item difficulty, they should usually remain matched on reliability for a normal group of a different mean ability level. The match should be on mean, variance, and shape of the distribution of item difficulty. Comparable shapes of the distributions of test scores are also necessary if scores from the two tests are to be meaningfully compared. The matching of the two tests using normal subjects of a wide range of ability will usually result in the two tests retaining equivalent reliabilities for normal subgroups of different levels of ability. This equivalence must be verified for any particular group of pathological subjects by recomputing test characteristics for a normal sample which scores at the level of the pathological group. With such equivalence any difference between the two test scores for pathological subjects can be attributed to genuine differential deficit in ability.

Despite fairly close matching on the above variables, tests may differ slightly on mean or variance for normal subjects. Tests may also differ slightly on number of items. If, in such cases, the two distributions of scores have similar shapes, one might reasonably convert all scores to standard scores on the basis of the performance of the normal standardization sample. This conversion corrects for differences between the tests in mean and variance and facilitates comparison. (It does not, of course, correct for differences between tests in discriminating power.)

Illustrative Example

The need for equivalent discriminating power may be illustrated by some data recently gathered by the present writers. A 114-item multiple-choice analogies test with items of gradated difficulty was given to 49 severely disturbed schizophrenics and 206 normal subjects. This test was used as a pool of items from which various subsets of items were drawn to compare schizophrenic and normal performance for purposes of this paper. The

114 items were arbitrarily divided to yield two forms. Form A consisted of the 62 items for which the first word of the analogy began with a letter that occurs in the alphabet between A and L. Form M consisted of the remaining 52 items for which the first letter of the word fell between M and Z. Normal subjects scored about the same on the two forms, and the schizophrenics showed about the same performance deficit on one form as on the other.

The effects of difference in reliability on performance deficit may be illustrated by drawing subtests from the two forms with identical means but different reliabilities for normal subjects. Coefficient alpha (Kuder-Richardson Formula 20) was our measure of reliability. As shown in Table 1, a pair of subtests was drawn with a coefficient alpha of .80, one subtest from each form. A second pair of subtests was drawn with lower coefficient alpha values, .47 for the subtest from Form A and .49 for the subtest from Form M. These four subtests were closely matched on mean accuracy. The more reliable subtests had the lower variances of item difficulty and the higher test variances, as might be expected. Table 1 shows the mean accuracy of the schizophrenic and normal subjects for each subtest. The schizophrenics showed, as predicted, a greater preformance deficit on the subtests with the higher reliabilities, whether from Form A or M. Thus, a comparison of the schizophrenic performance on the *less* reliable Form M subtest with the *more* reliable Form A subtest yields poorer schizophrenic accuracy on Form A than on Form M ($t = 4.89$, $p < .001$). On the other hand, if one should choose the *more* reliable Form M subtest and the *less* reliable Form A subtest, one would obtain the opposite finding, with the schizophrenics scoring lower on Form M than on Form A ($t = 4.32$, $p < .001$). These results show how easily differences in reliability may artifactually yield an apparent differential deficit in ability. If the two forms were designed to measure different abilities, as in the usual study of schizophrenic cognition, the investigator could easily conclude from performance on either one of these pairs of subtests that he had demonstrated a differential deficit in ability.

Let us now examine data that show how differential difficulty of items can produce differences in discriminating power of tests despite equivalent reliabilities (both difficulty and reliability again measured by normal performance alone). These data are from a free response vocabulary test that was given to 52 schizophrenics and 56 normal control subjects, some of the items being affect laden (emotional) and others being affectively neutral. Table 2 shows the performance of schizophrenic and normal subjects on two 10-item emotional and two 10-item neutral subtests with the coefficient alphas around .70. Two subtests, one emotional and one neutral, were chosen from items of medium difficulty and two others, one emotional and one neutral, from very difficult items. The items of medium difficulty were of about 65% accuracy for the normal subjects alone and about 51% for the two groups combined. The very difficult items were of less than 10% accuracy for

TABLE 1

RELATIONSHIP OF SCHIZOPHRENICS' PERFORMANCE DEFICIT TO TEST RELIABILITY

Subtest	Number of items	Coefficient alpha	Variance of item difficulty	Accuracy			
				Normals		Schizophrenics	
				$\bar{X}$	*SD*	$\bar{X}$	*SD*
High reliability							
Form A	10	.80	.02	6.3	2.8	3.0	2.8
Form M	10	.80	.03	6.4	2.7	3.0	2.6
Low reliability							
Form A	10	.47	.08	6.5	1.6	4.2	1.9
Form M	10	.49	.08	6.3	1.7	4.6	2.1

TABLE 2

Relationship of Schizophrenics' Performance Deficit to Test Difficulty

Degree of difficulty	Number of items	Coefficient alpha	Variance of item difficulty	Accuracy			
				Normals		Schizophrenics	
				$\bar{X}$	*SD*	$\bar{X}$	*SD*
Medium							
Emotional	10	.68	.04	6.4	2.3	3.8	3.2
Neutral	10	.68	.04	6.6	2.2	3.8	2.8
High							
Emotional	10	.73	.003	.7	1.3	.8	1.7
Neutral	10	.71	.003	.7	1.3	.5	1.2

each group. The schizophrenics, as expected, did not differ from the normal subjects on the very difficult subtests ($t = .32$ for emotional and .72 for neutral), but showed a larger deficit on the easier subtests ($t = 4.76$ for emotional and 5.84 for neutral; $p < .001$ for both). The investigator who compared groups on one emotional and one neutral subtest might conclude that he had found a greater schizophrenic deficit in emotional than in neutral vocabulary, or vice versa, if he chose subtests of differing levels of difficulty.

Alternative Ways to Rule Out Differential Discriminating Power as an Artifact

In some situations, tasks unmatched on discriminating power may give legitimate evidence of differential deficit in ability. If the control task is a more discriminating measure of nonpathological differences in ability than the experimental task, and if, despite this disparity, the experimental task yields the greater difference between pathological and normal performance, one must attribute the differential performance to a true differential deficit rather than to a generalized deficit coupled with a difference between the tasks on discriminating power. For example, the more reliable subtests of standard intelligence tests are usually more discriminating than the less carefully constructed experimental tasks that are used in most studies of pathological deficit.

A less adequate approach that is sometimes used is to select atypically poor-scoring control subjects so as to match pathological and normal groups on accuracy on the control task. This method is often complicated by the problems of statistical regression. One must select atypically good-scoring pathological subjects and atypically poor-scoring normal subjects on the control task to achieve the matching, and these atypically scoring subgroups usually regress toward the means of the larger groups on their scores on the experimental task. Matching subgroups may also produce systematic unmatching on other variables, as recently discussed by Meehl (1971). The usefulness of findings from a study using such matching ranges from fair to zero depending on how extremely atypical the subjects must be to achieve the matching.

Statistical Solutions

Several investigators have attempted statistical solutions to the problem of demonstrating differential deficit with tests of differing reliabilities. These statistical solutions include (*a*) analysis of covariance to covary out the scores on the control task from scores on the experimental task, or regression analysis to estimate scores on the experimental task from score on the control task, and then assign a residualized error score to each subject, and (*b*) estimation of error-free true scores by correcting obtained scores using measures of reliability. Neither of these methods is adequate. In each case the more discriminating test will still yield a greater performance deficit than the less discriminating test when given to a group of subjects with lower ability than the control group. A statistical solution to the problem may be possible, but we are not aware that one has yet been developed.

REFERENCES

LORD, F. M. The relation of the reliability of multiple-choice tests to the distribution of item difficulties. *Psychometrika,* 1952, **17**, 181–194. (a)

LORD, F. M. A theory of test scores. *Psychometric Monographs,* 1952, No. 7. (b)

MEEHL, P. E. High school yearbooks: A reply to Schwarz. *Journal of Abnormal Psychology,* 1971, **77**, 143–148.

TRUSCOTT, I. P. Contextual constraint and schizophrenic language. *Journal of Consulting and Clinical Psychology,* 1970, **35**, 189–194.

(Received May 22, 1972)

Psychological Bulletin
1973, Vol. 79, No. 6, 386–388

IS THERE ANY PROPRIOCEPTIVE FEEDBACK?

COMMENTS ON SCHMIDT (1971)

BILL JONES[1]

University of Queensland

Schmidt argued that experiments by himself and his co-workers support the theory that the timing of skilled movements is based upon proprioceptive feedback. The author argues that Schmidt's account of timing is too narrow to describe skilled performance and tries to show that the experiments Schmidt relies upon do not rule out an alternative theory of timing in terms of central monitoring of efferent signals to the muscles.

Any skilled task demands that the central nervous system have the ability to control the temporal ordering of a pattern of movements. Even a "discrete" voluntary movement requires at the least the coordinated regulation of two antagonistic muscle groups at a joint. This central capacity to order the temporal sequence of movements is ordinarily referred to as "timing" (Bartlet, 1947; Lashley, 1951). Schmidt (1971) has argued that timing depends upon feedback from proprioceptors in joints, tendons, and muscles. Schmidt, in practice, equates timing with timekeeping or "time estimation," and most of the experimental paradigms he reviews are concerned with such discrete acts as pressing a key or switch. However, his paper implies that he wishes to place his proprioceptive feedback theory in the wider context of skill (i.e., any *sequence* of movements which meets some criterion of correctness). For example, he argues (p. 383) that the high levels of skill displayed by well-practiced subjects depend upon the timing (i.e., temporal ordering) of the various submovements, and in contrasting his feedback or input theory with a preprogramming hypothesis, Schmidt says that the preprogram "dictates the timing of the various submovements in a sequence [p. 392]."

In the experiments discussed by Schmidt, the subject's task was normally to respond to questions such as, *At which time shall I press the key*, which required him to estimate a period of time given, it is assumed, proprioceptive cueing. In contrast, consider any transcription skill such as typewriting where the subject's task is not to estimate the passage of time in order to control the sequence of movements. Since he has not merely to press any key but to press the correct key in the sequence, the subject needs to know the extent and direction of previous submovements in the pattern. In short, the skill requires accurate placement of the fingers, and information as to position of the fingers is a prerequisite of timing. It is, of course, a debatable point, whether such information comes from proprioceptive feedback or through some monitoring by the central nervous system of its own motor outflow (i.e., preprogramming) and in a brief note, it is hardly possible to review the evidence for either position. It may simply be noted here that there is strong evidence which would be difficult to reconcile with any existing proprioceptive input theory (see e.g., Angel, Garland, & Fischler, 1971; Higgins & Angel, 1970; Jones, 1972).

If we adopt Schmidt's restricted usage of timing as time estimation, it does not seem to me that the experiments he discusses rule out the central monitoring hypothesis as opposed to proprioceptive feedback model. In a task which required subjects to estimate a short time interval by moving a slide for the correct duration over a given distance, Ellis (1969) found that absolute error of time estimation decreased when acceleration, velocity, and tension cues were manipulated compared to no-resistance controls, being most accurate for manipulations of acceleration. Further, subjects were required to smell two-, three-, and four-letter words with auditory cues blocked by white noise, and time estimation was more accurate with four-letter words. When spelling was withdrawn, there was a marked decline in accuracy of time estimation for the no-resistance controls with only slight decreases for the resistance groups.

---

[1] Requests for reprints should be sent to Bill Jones, Department of Psychology, University of Queensland, St. Lucia, Queensland, Australia 4067.

Schmidt (1971) argued that spelling provided proprioceptive feedback from the larynx, lips, and tongue. However, it may equally well have provided a self-pacing strategy via *motor outflow* to the larynx, etc., since, of course, if spelling is discontinued, not only is there an end to feedback signals from the larynx but there is also an end to motor outflow to the larynx. It is perhaps worth noticing that proprioceptive signals from the larynx do not provide the singer with sufficient information to control pitch since accuracy of pitching declines when auditory feedback is withdrawn (Deutsch & Clarkson, 1959). It may be argued that Deutsch and Clarkson are dealing with continuous graded activity whereas Schmidt is arguing about a discrete response. However, the so-called discrete response still involves temporal patterning of muscular contractions and may therefore be compared with "continuous" responses.

The results in the study by Schmidt and Christina (1969) did, in general, indicate that with higher levels of movement (Schmidt begs the question by saying proprioceptive feedback) provided through rotation of a lever, the subject's ability to anticipate the time of arrival of a signal was facilitated. The subject was required to rotate a lever with the left hand over a time course of 1.5 seconds, the time taken by a moving pointer, which was started by the subject's action on the lever, to reach a stationary target. The subject's main task was to release a reaction-time key with the right hand when pointer and target became aligned.

There were three feedback conditions: intermediate and high, in which the subject rotated the lever through 3.5 inches and 11.5 inches, respectively, and minimal, where the subject merely depressed a switch with the left hand to activate the moving pointer and then rested his arm on his left knee. In the main, lower reaction times were obtained with the higher levels of movement. Proprioceptive-feedback manipulations may, therefore, be important in increasing accuracy of velocity control and so provide a cue for time estimation (but that is not necessary evidence that proprioceptors are involved in the control of limb placement within temporal patterns).

On the other hand, since subjects made voluntary movements there must have been signals from the central nervous system to the muscles which could have been centrally monitored. Schmidt and Christina's results may only indicate that such monitoring is facilitated by increasing the amount of movement involved. The central theory could only be ruled out if Schmidt and Christina's results were replicable with passive rather than active carriage of the left arm. Since subjects were, in fact, free to move as far as they pleased, the end point of each movement could have been preprogrammed, that is, determined by patterns of outflow within the central nervous system rather than proprioceptive inflow and the right-hand response cued on a central rather than peripheral basis (see Jones, 1972).

To summarize, Schmidt (1971) implies that studies of timekeeping throw light on the coordination of skilled movements and, more specifically, that proprioceptive feedback is involved in timekeeping. However, it is clear that the coordination of a temporal sequence of movements in a skill requires much more than time estimation; it requires knowledge of direction and extent of previous movements in the sequence since movements must be accurate. Second, the studies which Schmidt (1971) regards as central to the proprioceptive feedback theory of timekeeping (Christina, 1970; Ellis, 1969; Schmidt & Christina, 1969) do not, in fact, provide evidence that proprioception is necessarily involved in time estimation since if subjects make voluntary movements, there must be central nervous system outflow which could serve as an alternative basis for time estimation.

REFERENCES

Angel, R., Garland, M., & Fischler, M. Tracking errors amended without visual feedback. *Journal of Experimental Psychology*, 1971, **89**, 422–424.

Bartlett, F. C. The measurement of human skill. *British Medical Journal*, 1947, **1**, 835–858, 877–880.

Christina, R. W. Proprioception as a basis for the temporal anticipation of motor responses. *Journal of Motor Behavior*, 1970, **2**, 125–133.

Deutsch, A., & Clarkson, T. K. Nature of the vibrato and the control loop in singing. *Nature*, 1959, **183**, 167–168.

Ellis, M. J. Control dynamics and timing a discrete motor response. *Journal of Motor Behavior*, 1969, **1**, 119–134.

Higgins, J. R., & Angel, R. W. Correction of tracking errors without sensory feedback. *Journal of Experimental Psychology*, 1970, **84**, 412–416.

Jones, B. Outflow and inflow in movement duplication. *Perception and Psychophysics*, 1972, **12**, 95–96.

Lashley, K. S. The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior: The Hixon symposium.* New York: Wiley, 1951.

Schmidt, R. A. Proprioception and the timing of motor responses. *Psychological Bulletin*, 1971, **76**, 383–393.

Schmidt, R. A., & Christina, R. W. Proprioception as a mediator in the timing of motor responses. *Journal of Experimental Psychology*, 1969, **81**, 303–307.

(Received June 7, 1972)

Psychological Bulletin
1973, Vol. 79, No. 6, 389–390

PROPRIOCEPTION VERSUS MOTOR OUTFLOW IN TIMING:

A REPLY TO JONES

RICHARD A. SCHMIDT

University of Michigan

Jones' point that the motor outflow hypothesis should be considered a rival to the proprioceptive feedback interpretation of the motor response timing data reviewed by Schmidt has potential merit. However, with no data ruling in favor of motor outflow over proprioceptive feedback for timing, and with other peripheral feedback channels (e.g., vision) known to be mechanisms for timing, the proprioceptive feedback explanation is favored over motor outflow.

Rather than the proprioceptive feedback interpretation for some of the data I reviewed in my paper (Schmidt, 1971), Jones (1973) prefers an interpretation wherein motor outflow to the muscles provides a basis for the timing of future responses. Jones' point is that the data discussed in terms of the proprioceptive feedback input hypothesis did not rule out the possibility of motor outflow, and that, therefore, the motor outflow view was also deserving of attention as a possible explanation for that body of data. Strictly speaking, he is correct.

While space does not permit an extensive comparison of the motor outflow and proprioceptive feedback hypotheses for movement control, it should be pointed out that, using motor responses, there seem to be two lines of evidence for the outflow hypothesis which cannot be handled by other theories. First is the evidence by Angel, Garland, and Fischler (1971) who showed that subjects could correct errors in step tracking without peripheral feedback. Second, the evidence by Taub and Berman (1968) that monkeys could *learn* motor responses after deafferentation may indicate that some form of feedback was mediating learning. These studies point to the possibility of internal monitoring of the motor outflow to the muscles via a rapid feedback loop as a mechanism in motor learning and performance. To the contrary, that a subject can execute movement in the absence of any peripheral feedback (e.g., Lashley, 1917) and that the accuracy of duplication is a function of active versus passive presentation (e.g., Jones, 1972) can be handled easily by the notion of the motor program (see Keele, 1968; Schmidt, 1972; Schmidt & Russell, 1972). The problem with all this evidence for the present question is that it deals with the *spatial* aspects of movement production and error correction, whereas the present issue concerns mechanisms for determining the *time* of initiation of these sequences (e.g., the time to initiate a swing so as to hit a pitched ball). Of course, the possibility of internal monitoring for one class of behavior does not provide assurance that internal monitoring is involved in timing.

In sharp contrast to the suggestion that outflow is involved, there is strong evidence that peripheral information plays a powerful role in timing. Consider vision, for example, in a situation where the subject must respond exactly 2.0 seconds after a starting signal while either watching or not watching a clock. Some of our unpublished evidence shows that with vision, subjects make absolute errors of only a few milliseconds, while without vision (e.g., Quesada & Schmidt, 1970) the mean absolute error is about 100 milliseconds. Clearly inflow from the periphery can aid timing. Since there is no logical necessity to assume any fundamental difference between proprioceptive feedback and the other senses (other than, of course, those differences involving the specific receptors or the types of physical stimuli detected), it would seem that Jones' (1973, p. 5) suggestion that the proprioceptive feedback hypothesis could be ruled out if passive movements had been employed in the Schmidt-Christina (1969) experiment

---

[1] Reprint requests should be sent to Richard A. Schmidt, Department of Physical Education, University of Michigan, Ann Arbor, Michigan 48104.

would be certain to support the proprioceptive feedback hypothesis. What is the difference between having a subject watch a clock or having his limb moved predictably for 2.0 seconds? Both provide reliable sensory information about the passage of time, and proprioceptive feedback would almost certainly be superior to no feedback at all. In fact, we (Schmidt & Christina, 1969) did not use passive movements because we wanted to rule out this explanation, as we were interested *not* in whether the subject *could* use proprioceptive feedback to time (which seemed obvious) but whether, in fact, he *did* use the feedback. Our findings indicated that he did, although we must admit the possibility that they could have been using motor outflow information.

Therefore, since there is the possibility that a motor outflow mechanism may be involved in timing as it may be with motor control situations, and since the timing data do not strictly rule out this view, Jones' point that it should be considered a rival to the proprioceptive feedback hypothesis is well taken. However, because there is support that peripheral information is strongly involved in timing, and there is no direct support that a motor outflow mechanism is involved, I naturally prefer to interpret the timing data in proprioceptive feedback terms. The answer to Jones's question, "Is there any proprioceptive feedback?" seems to be "Yes" for timing. We now anxiously await solid data which answer affirmatively the question, "Is there any outflow in timing?"

REFERENCES

Angel, R. W., Garland, H., & Fischler, M. Tracking errors amended without visual feedback. *Journal of Experimental Psychology*, 1971, **89**, 422–424.

Jones, B. Outflow and inflow in movement duplication. *Perception and Psychophysics*, 1972, **12**, 95–96.

Jones, B. Is there any proprioceptive feedback? Comments on Schmidt (1971). *Psychological Bulletin*, 1973, **79**, 386–390.

Keele, S. W. Movement control in skilled motor performance. *Psychological Bulletin*, 1968, **70**, 387–403.

Lashley, K. S. The accuracy of movement in the absence of excitation from the moving organ. *American Journal of Physiology*, 1917, **43**, 169–194.

Quesada, D. C., & Schmidt, R. A. A test of the Adams-Creamer decay hypothesis for the timing of motor responses. *Journal of Motor Behavior*, 1970, **2**, 273–283.

Schmidt, R. A. Proprioception and the timing of motor responses. *Psychological Bulletin*, 1971, **76**, 383–393.

Schmidt, R. A. The index of preprogramming (IP): A statistical method for evaluating the role of feedback in simple movements. *Psychonomic Science*, 1972, **27**, 83–85.

Schmidt, R. A., & Christina, R. W. Proprioception as a mediator in the timing of motor responses. *Journal of Experimental Psychology*, 1969, **81**, 303–307.

Schmidt, R. A., & Russell, D. G. Movement velocity and movement time as determiners of the degree of preprogramming in simple movements. *Journal of Experimental Psychology*, 1972, **96**, 315–320.

Taub, E., & Berman, A. J. Movement and learning in the absence of sensory feedback. In S. J. Freedman (Ed.), *The neuropsychology of spatially oriented behavior*. Homewood, Ill.: Dorsey, 1968.

(Received June 7, 1972)

Psychological Bulletin
1973, Vol. 79, No. 6, 391–395

COMPUTATIONAL PROCEDURES FOR ESTIMATING MAGNITUDE OF EFFECT FOR SOME ANALYSIS OF VARIANCE DESIGNS

DAVID H. DODD[1] AND ROGER F. SCHULTZ, JR.
University of Utah

Computational procedures are presented for determining estimates of magnitude of effect or proportion of variance (ω^2) for a variety of analysis of variance designs. Tables summarize simplified computational formulas for fixed, random, and mixed designs, including both nonrepeated- and repeated-measures cases.

Vaughan and Corballis (1969) have discussed the problem of estimating magnitude of effect in various analysis of variance (ANOVA) designs. Their paper emphasized the importance of reporting measures of degree of association between the independent variable(s) and the dependent variable. Most significantly, it makes available a general technique for finding such estimates, a technique based directly on expected mean squares. The technique can be applied to virtually any ANOVA model, although certain designs do not, in general, allow unbiased estimates.

The primary purpose of this article is to provide simple computational procedures for some frequently encountered designs. All of the formulas are derived directly from the expected mean squares, E(MS)s, for each source of variance. Some algebra yields formulas which are generally simpler than those presented by Vaughan and Corballis. In addition, they eliminate the need to determine separately an estimate of variance for every source. This computation can be avoided where a variance source is not of interest to the investigator or is small enough, relatively speaking, to be ignored. Another purpose of this paper is to present procedures for estimating magnitude of effect in some of the more common repeated-measures designs, many of which are not presented in Vaughan and Corballis.

The notation used here is basically that of Vaughan and Corballis. Each factor in a design is symbolized by an Arabic letter. Fixed and random factors are indicated by capital and lowercase letters, respectively. To indicate which factor(s) is involved in repeated measurement, overlining is employed, for example, $\bar{A}$. Since subjects (a random factor) always becomes a factor in repeated-measures designs, the letter S is included in the symbol for such designs, for example, $\bar{A}BS$ is a design having Fixed Factors A and B and repeated measurement on A.

FIXED FACTORS HAVING NO REPEATED MEASURES

The computational formula for these designs is given by Hays (1963). Regardless of the number of factors, estimates are obtained by essentially the same formula.

In the one-way ANOVA, the proportion of variance accounted for by the treatments is of interest:

$$\hat{\omega}^2_{\text{between}} = \frac{SS_{\text{between}} - (df_{\text{between}} \times MS_{\text{within}})}{SS_{\text{total}} + MS_{\text{within}}}.$$

The result is essentially the same for the two-way ANOVA. This case is used to illustrate the entire derivation; consider Factor A. The numerator consists of the variance estimate for A given in Table 1. The denominator consists of the sum of all variance estimates. Multiplying numerator and denominator by nJK and simplifying yields the results indicated. If the F for a factor is relatively low, or the factor is not of interest, that calculation can be omitted.

In general, for any size ANOVA involving fixed factors where none are repeated, the computational formula is

$$\hat{\omega}^2_{\text{effect}} = \frac{SS_{\text{effect}} - (df_{\text{effect}} \times MS_{\text{error}})}{SS_{\text{total}} + MS_{\text{error}}}.$$

---

[1] Requests for reprints should be sent to David H. Dodd, Department of Psychology, University of Utah, Salt Lake City, Utah 84112.

TABLE 1

EXPECTED MEAN SQUARES, ESTIMATES OF VARIANCE, AND NUMERATORS FOR ESTIMATING MAGNITUDE OF EFFECT FOR THE COMPONENTS OF A TWO-WAY FIXED FACTOR DESIGN

Source	E(*MS*)	Estimate of variance	Numerator for computational procedure
A	$\sigma_e^2 + nK\theta_A^2 = \sigma_e^2 + \frac{nJK}{J-1}\sigma_A^2$	$\frac{(J-1)(MS_A - MS_{\text{error}})}{nJK}$	$SS_A - (df_A \times MS_{\text{error}})$
B	$\sigma_e^2 + nJ\theta_B^2 = \sigma_e^2 + \frac{nJK}{K-1}\sigma_B^2$	$\frac{(K-1)(MS_B - MS_{\text{error}})}{nJK}$	$SS_B - (df_B \times MS_{\text{error}})$
AB	$\sigma_e^2 + n\theta_{AB}^2 = \sigma_e^2 + \frac{nJK}{(J-1)(K-1)}\sigma_{AB}^2$	$\frac{(J-1)(K-1)(MS_{AB} - MS_{\text{error}})}{nJK}$	$SS_{AB} - (df_{AB} \times MS_{\text{error}})$
Error	σ_e^2	MS_{error}	

Note. The denominator for all estimates of magnitude of effect is $SS_{\text{total}} + MS_{\text{error}}$.

FIXED MODEL WITH REPEATED MEASURES

For designs in which there is repeated measurement on a fixed factor, there is some degree of underestimation in estimates of proportion of variance since the expected value of the denominator is somewhat greater than the sum of the true variances. In many cases, the investigator may simply choose to accept this bias, which may be relatively small. The

TABLE 2

FIXED FACTOR REPEATED-MEASURE DESIGN NONADDITIVE AND ADDITIVE MODELS

Design	Conditions	Denominator	Degree of bias in denominator
$\bar{A}S$			
Nonadditive	None	$SS_{\text{total}} + MS_S + nMS_{A\times S}$	$+\sigma_e^2/J$
Additive	No $A \times S$ effect	$SS_{\text{total}} + MS_S$	None
$A\bar{B}S$			
Nonadditive	None	$SS_{\text{total}} + MS_S + nJMS_{B\times S}$	$+\sigma_e^2/K$
Additive	No $B \times S$ effect	$SS_{\text{total}} + MS_S$	None
$\bar{A}\bar{B}S$			
Nonadditive	None	$SS_{\text{total}} + MS_S + nMS_{A\times S} + nMS_{B\times S} + n(J+K-1)MS_{AB\times S}$	$\frac{\sigma_e^2}{JK} + \frac{\sigma_e^2}{J} + \frac{\sigma_e^2}{K}$
Additive	(*a*) No $AB \times S$ effect	$SS_{\text{total}} + MS_S + nMS_{A\times S} + nMS_{B\times S} - 2nMS_{AB\times S}$	None
	(*b*) No $A \times S$, $B \times S$, or $AB \times S$ effects[a]	$SS_{\text{total}} + MS_S$	None
$AB\bar{C}S$			
Nonadditive	None	$SS_{\text{total}} + MS_S + nJKMS_{C\times S}$	$+\sigma_e^2/L$
Additive	No $C \times S$ effect	$SS_{\text{total}} + MS_S$	None
$A\bar{B}\bar{C}S$			
Nonadditive	None	$SS_{\text{total}} + MS_S + nJMS_{B\times S} + nJMS_{C\times S} + nJ(K+L-1)MS_{BC\times S}$	$\frac{\sigma_e^2}{KL} + \frac{\sigma_e^2}{K} + \frac{\sigma_e^2}{L}$
Additive	(*a*) No $BC \times S$ effect	$SS_{\text{total}} + MS_S + nJMS_{B\times S} + nJMS_{C\times S} - 2nJMS_{BC\times S}$	None
	(*b*) No $B \times S$, $C \times S$, or $BC \times S$ effects[b]	$SS_{\text{total}} + MS_S$	None

[a] Pooled $MS_{\text{error}} = \frac{SS_{A\times S} + SS_{B\times S} + SS_{AB\times S}}{df_{A\times S} + df_{B\times S} + df_{AB\times S}}$. The pooled MS_{error} is used as the denominator for all F ratios and appears in the numerator of all ω^2, for example, $SS_{AB} - (J-1)(K-1)MS_{\text{error}}$.

[b] Pooled $MS_{\text{error}} = \frac{SS_{B\times S} + SS_{C\times S} + SS_{BC\times S}}{df_{B\times S} + df_{C\times S} + df_{BC\times S}}$. The pooled MS_{error} is used as the denominator for all F ratios (except for A which uses MS_S) and thus appears in the numerator for all ω^2 (except A), for example, $SS_{BC} - (K-1)(L-1)MS_{\text{error}}$.

Note. Numerators are SS effect $- (df_{\text{effect}} \times MS_{\text{appropriate error term}})$ for the effects of A, B, C, and their interactions.

direction of bias is such that the investigator knows that his effect is at least as strong as the obtained estimate. However, as suggested by Vaughan and Corballis, the additive model can be assumed, that is, one can assume that there are no interactions of subjects with repeated-measures factors. A test of this assumption was devised by Tukey (1949) and is described by Winer (1962, p. 217), Myers (1966, p. 166), and Kirk (1968, p. 137). Unfornately, this test is both laborious to calculate and restricted to the assessment of possible Linear × Linear interaction components. A second solution is sometimes possible. If, for each subject, replications on at least one of the repeated-measures factors can be obtained as part of the experiment, then an independent estimate of error variance can be made, allowing unbiased estimates of magnitude of effect. This is not possible in any experimental settings where secular trends might be expected, that is, where changes in the dependent variable are associated with the time or trial of testing. For example, secular trends are expected when learning trials are the repeated-measures factor. In some kinds of research (e.g., psychophysics), it is both possible and even efficient to obtain more measures from each trained subject. The variance between the replicates (measures on the same subject under a given treatment combination) estimates error variance. With replications, the repeated-measures, fixed-effect design becomes identical to the mixed model; the subjects' variable is a random-effects variable and replications are the within-treatments error component.

TABLE 3

COMPUTATIONAL COMPONENTS FOR ESTIMATES OF MAGNITUDE OF EFFECT FOR RANDOM AND MIXED DESIGNS

Design	Source	Numerator	Denominator
Random			
a	A	$J(MS_A - MS_S)$	$SS_{\text{total}} + MS_A$
ab	A	$J(MS_A - MS_{AB})$	$SS_{\text{total}} + MS_A + MS_B - MS_{AB}$
	B	$K(MS_B - MS_{AB})$	
	AB	$JK(MS_{AB} - MS_S)$	
abc	A	$J(MS_A - MS_{AB} - MS_{AC} + MS_{ABC})$	$SS_{\text{total}} + MS_A + MS_B + MS_C$ $- MS_{AB} - MS_{AC} - MS_{BC}$ $+ MS_{ABC}$
	B	$K(MS_B - MS_{AB} - MS_{BC} + MS_{ABC})$	
	C	$L(MS_C - MS_{AC} - MS_{BC} + MS_{ABC})$	
	AB	$JK(MS_{AB} - MS_{ABC})$	
	AC	$JL(MS_{AC} - MS_{ABC})$	
	BC	$KL(MS_{BC} - MS_{ABC})$	
	ABC	$JKL(MS_{ABC} - MS_S)$	
Mixed			
aB	A	$J(MS_A - MS_S)$	$SS_{\text{total}} + MS_A + JMS_{AB} - JMS_S$
	B	$SS_B - (df_B \times MS_{AB})$	
	AB	$JK(MS_{AB} - MS_S)$	
aBC	A	$J(MS_A - MS_S)$	$SS_{\text{total}} + MS_A + JMS_{AB} + JMS_{AC}$ $+ J(K + L - 1)MS_{ABC}$ $- J(K + L + 1)MS_S$
	B	$SS_B - (df_B \times MS_{AB})$	
	C	$SS_C - (df_C \times MS_{AC})$	
	AB	$JK(MS_{AB} - MS_S)$	
	AC	$JL(MS_{AC} - MS_S)$	
	BC	$SS_{BC} - (df_{BC} \times MS_{ABC})$	
	ABC	$JKL(MS_{ABC} - MS_S)$	
abC	A	$J(MS_A - MS_{AB})$	$SS_{\text{total}} + MS_A + MS_B - MS_{AB}$ $+ JMS_{AC} + JMS_{BC}$ $+ (JK - J - K)MS_{ABC}$ $- JKMS_S$
	B	$K(MS_B - MS_{AB})$	
	C	$SS_C - [df_C \times (MS_{AC} + MS_{BC} - MS_{ABC})]$	
	AB	$JK(MS_{AB} - MS_S)$	
	AC	$JL(MS_{AC} - MS_{ABC})$	
	BC	$KL(MS_{BC} - MS_{ABC})$	
	ABC	$JKL(MS_{ABC} - MS_S)$	

TABLE 4

COMPUTATIONAL COMPONENTS FOR RANDOM AND MIXED DESIGNS WITH REPEATED MEASURES

Design	Source	Numerator	Denominator
Random			
$\bar{a}S$		Solution identical to the *ab* design. See text.	
$\bar{a}\bar{b}S$		Solution identical to the *abc* design. See text.	
$a\bar{b}S$	A	$J(MS_A - MS_S - MS_{AB} + MS_{B\times S})$	$SS_{\text{total}} + MS_A + MS_B - MS_{AB}$
	B	$K(MS_B - MS_{AB})$	
	AB	$JK(MS_{AB} - MS_{B\times S})$	
	S	$nJ(MS_S - MS_{B\times S})$	
Mixed			
$A\bar{b}S$	A	$SS_A - [(df_A) \times (MS_S + MS_{AB} - MS_{B\times S})]$	$SS_{\text{total}} + MS_S + MS_B + KMS_{AB}$ $- (K+1)MS_{B\times S}$
	B	$K(MS_B - MS_{B\times S})$	
	AB	$JK(MS_{AB} - MS_{B\times S})$	
	S	$nJ(MS_S - MS_{B\times S})$	
$a\bar{B}S$	A	$J(MS_A - MS_S)$	(1) Nonadditive case[b]: $SS_{\text{total}} + MS_A + JMS_{AB}$ $+ J(n-1)MS_{B\times S}$ (2) Additive case (no $B \times S$ effect): $SS_{\text{total}} + MS_A + JMS_{AB}$ $- JMS_{B\times S}$
	B	$SS_B - (df_B \times MS_{AB})$	
	AB	$JK(MS_{AB} - MS_{B\times S})$	
	S[a]	$nJ(MS_S - MS_{B\times S})$	
$\bar{a}\bar{B}S$	A	$J(MS_A - MS_{A\times S})$	(1) Nonadditive case[b]: $SS_{\text{total}} + MS_S + MS_A - MS_{A\times S}$ $+ nMS_{B\times S} + JMS_{AB}$ $+ (nJ - n - J)MS_{AB\times S}$ (2) Additive case (no $AB \times S$): $SS_{\text{total}} + MS_S + MS_A + MS_{A\times S}$ $+ nMS_{B\times S} + JMS_{AB}$ $- (n+J)MS_{AB\times S}$
	B	$SS_B - df_B(MS_{B\times S} + MS_{AB} - MS_{AB\times S})$	
	AB	$JK(MS_{AB} - MS_{AB\times S})$	
	S	$n(MS_S - MS_{A\times S})$	

[a] Additive model only.

[b] Denominator too large by σ_e^2/K.

Table 2 presents the one-, two-, and three-factor designs. The results can be easily extended to four-factor and larger designs, where only one or two of the factors are repeated. The table also provides the necessary information as to the conditions for the additive model, the degree of bias in the denominator, etc. Note in particular that for all of these designs the numerator is given by

$$SS_{\text{effect}} - (df_{\text{effect}} \times MS_{\text{appropriate error term}}),$$

where the appropriate error term is the denominator of the F ratio for that effect.

RANDOM AND MIXED WITH NO REPEATED MEASURES

Consider, for example, a design having two random-effects variables. In contrast to $E(MS_A)$ for fixed effects in Table 1, the $E(MS_A)$ for the random model is:

$$\sigma_e^2 + nK\sigma_A^2.$$

Computational formulas are thus not quite like the fixed-effects results. In general, the numerators for estimates of random-effects variables are given by

$$(\text{Number of levels of factor}) \times (MS_{\text{effect}} - MS_{\text{appropriate error term}}).$$

The entire computational procedure for random and mixed designs is summarized in Table 3. These results can be readily generalized to larger designs.

RANDOM AND MIXED DESIGNS WITH REPEATED MEASURES

Solutions for these designs are more complicated. Whenever the repeated measurement is made on a fixed factor, the estimates show some degree of negative bias. Because of the large number of designs generated by three factors, Table 4 gives solutions for one- and two-factor designs only.[2] The $\bar{a}S$ and $\bar{a}\bar{b}S$ designs are, as noted, identical to the *ab* and *abc* designs, respectively. For the $\bar{a}S$ design, simply consider that S becomes b and n becomes K, then refer to Table 3. For the $\bar{a}\bar{b}S$ design, S becomes c and n becomes L.

Conclusions

It is hoped that the availability of tables giving direct computational formulas for the most common designs will encourage researchers to include estimates of magnitude of effects. Researchers are encouraged to read Vaughan and Corballis (1969) for a more complete discussion of a variety of issues and for the solution of special problems such as specific comparisons between treatment levels.

REFERENCES

Hays, W. L. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963.

Kirk, R. E. *Experimental design: Procedures for the behavioral sciences.* Belmont, Calif.: Brooks/Cole, 1968.

Myers, J. L. *Fundamentals of experimental design.* Boston: Allyn & Bacon, 1966.

Tukey, J. W. One degree of freedom for nonadditivity. *Biometrics,* 1949, 5, 232–242.

Vaughan, G. M., & Corballis, M. C. Beyond tests of significance: Estimating strength of effects in selected ANOVA designs. *Psychological Bulletin,* 1969, 72, 204–213.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received June 27, 1972)

---

[2] Tables covering the three-factor random and mixed repeated-measures designs can be obtained by writing to the first author.

Psychological Bulletin
1973, Vol. 79, No. 6, 396–397

DIRECT VISUAL PERCEPTION:

A REPLY TO GYR

JAMES J. GIBSON[1]

Cornell University

In his criticism of this author's theory of perception, Gyr did not take into account the basic issue of the availability of stimulus information and did not consider the radically new conception of proprioception and self-awareness that is an essential part of the theory.

The paper in the April 1972 *Psychological Bulletin* by John W. Gyr (1972) entitled "Is a theory of direct visual perception adequate?" does not, it seems to me, get at the deep issues. He objects to my theory of perception as published in *The Senses Considered as Perceptual Systems* (Gibson, 1966) and most recently in "Outline of a theory of direct visual perception" (Gibson, 1972). His grounds for objection are that I leave out of account the internal contribution to the perception of the external environment, or the fact that perception involves the construction of an internal model of the environment. (This kind of objection can take many forms and I think I am aware of all of them.) But Gyr also disagrees with my theory of proprioception, since he follows the ideas of von Holst as to how the observer gets information about his own activity. He seems to believe that this information about action or about plans for action provides the internal contribution to perception, and this I cannot understand. The problem of the adequacy of external information is one thing, and the problem of how to conceive proprioception and self-awareness is another.

First, as to the problem of mediated or direct perception, Gyr failed to consider the basic issue: whether there is *no* information in ambient light to specify the features of the world, as classical optics has taught, or whether there *is*. Perception can only be direct in the latter case. If the specifying invariants are normally available, an active observer can extract them and does not have to construct them. If they are *not* available, even under favorable circumstances, then the features of the world always have to be constructed in the mind in some way that we have not yet been able to agree upon. The process of extracting invariants from ambient light is not understood either, but it is so radically different from a process of constructing an internal model of the world that it would put the study of perception on an entirely new footing if we accepted the possibility. This is a genuine issue. It is *preliminary* to the study of the act of perception. It concerns the very nature of the environment to be perceived. It raises deep questions having to do with what philosophers call ontology and epistemology, but psychologists neglect them at their peril.

Next, there are all the problems arising from von Holst's theory of the "efference copy." Many investigators have recently taken them up, including Held and Festinger and their collaborators. They are century-old unsolved problems on what used to be called "sensations of innervation" and "ideo-motor action." But these are legacies of mentalism. There is nothing really new in the idea of what Gyr calls the organism "getting information about its own voluntary activity." He summarizes many experiments which show in a general way that perception and action cannot be separated, but that does not help to resolve the issue. All these experimenters presuppose classical sense impressions as the basis of perception. If sensory signals are barren to begin with, there is no use adding on motor commands. Anyone who considers the senses to be channels of sensation *has* to be a mentalist when it comes to perception. That is, he has to assume a mind that can copy, store, compare, match, decide, and issue commands—a man in the brain.

---

[1] Reprint requests should be sent to James J. Gibson, Department of Psychology, Uris Hall, Cornell University, Ithaca, New York 14850.

The whole idea of sensory signals and motor commands is wrong, together with the psychology that goes with it. The brain does not receive messages nor does it issue orders. Even the use of the terms *sensory* and *motor* is mentalistic. The concept of a perceptual system is in sharp contrast to the old notion of a sensory channel inasmuch as the brain takes its proper place as part of the system and is no longer the seat of the mind. A perceptual system is supposed to seek an equilibrium and reach an optimum, as a system should. This formula avoids both mentalism and mechanism whereas formulae like that of Gyr are simply combinations of mentalism and mechanism.

The theory of perceptual systems instead of senses puts an entirely new meaning on the terms proprioception and self-awareness. It implies a rejection of the distinction between proprioceptors and exteroceptors together with the doctrine of specific qualities of sensory nerves and the theory of sense data. It is upsetting, but it eliminates a vast area of muddle and confusion in psychology. It assumes that there is always available to every perceptual system both propriospecific information and exterospecific information at the same time. Thus one cannot be aware of the world without being potentially aware of the body, and *vice versa*. Each pole of experience implies the other. If this is true, all the experiments and observations on proprioception and kinesthesis for a century will have to be reinterpreted, for the terms will not mean what they they used to mean. A whole set of current experiments and controversies will go by the board if the modalities of sense are recognized as being unimportant for the activity of perception, for the theory of direct perception implies an equally direct awareness of the body of the observer and the adjustments of its perceptual organs. New experiments will have to be designed to test this theory.

REFERENCES

GIBSON, J. J. *The senses considered as perceptual systems.* Boston: Houghton Mifflin, 1966.

GIBSON, J. J. Outline of a theory of direct visual perception. In J. R. Royce & W. W. Rozeboom (Eds.), *The psychology of knowing.* New York: Gordon & Breach, 1972.

GYR, J. W. Is a theory of direct visual perception adequate? *Psychological Bulletin,* 1972, 77, 246–261.

(Received June 28, 1972)

Psychological Bulletin
1973, Vol. 79, No. 6, 398-407

LATENT INHIBITION[1]

ROBERT E. LUBOW[2]
Tel-Aviv University, Ramat-Aviv, Israel

A review of the latent inhibition literature, concerning the decremental effects of nonreinforced preexposure to the to-be-conditional stimulus on subsequent learning, is presented. Latent inhibition is found to be a broadly based phenomenon appearing across a variety of species: goldfish, goat, sheep, rat, rabbit, dog, human child, and, under special conditions, the human adult, and across a variety of tasks: classical conditioning, avoidance conditioning, Ivanov–Smolensky conditioning, conditioned emotional response, go/no-go discrimination, reaction time, and conditioned taste aversion paradigms. The stability of latent inhibition as well as its stimulus specificity and the effects of number of nonreinforced preexposures are examined. Current explanations of latent inhibition which include the habituation of the orienting response, selective filtering, specific antagonistic and complementary responses, and conditioned inhibition are discussed. The need for a combined learning and attention theory is suggested.

Latent inhibition has been defined as a decrement in learning performance which results from the nonreinforced preexposure of the to-be-conditioned stimulus (Lubow & Moore, 1959). Simply, the conditioned stimulus presented by itself, before being paired with the unconditioned stimulus, retards the subsequent development of a conditioned response when the conditioned stimulus and unconditioned stimulus are paired. The purpose of this paper is to describe the empirical base which substantiates the existence of latest inhibition, in particular, its generality across different species and learning tasks, and to identify parameters which affect latent inhibition.

SCOPE OF LATENT INHIBITION PHENOMENON

Distribution across Species

Latent inhibition has been demonstrated in goldfish (Braud, 1971), goats and sheep (Lubow, 1965; Lubow & Moore, 1959), rabbits (Lubow, Markman, & Allen, 1968; Reiss & Wagner, 1972; Siegel, 1969a, 1969b, 1970, 1971; Siegel & Domjan, 1971, Experiment 2), dogs (Konorski & Szwejkowska, 1952) and, of course, rats (Ackil & Mellgren, 1968); Ackil, Mellgren, Halgren, & Frommer, 1969; Anderson, Merrill, Dexter, & Alleman, 1968; Anderson, O'Farrell, Formica, & Caponigra, 1969a; Anderson, Wolf, & Sullivan, 1969b; Carlton & Vogel, 1967; Chacto & Lubow, 1967; Domjan & Siegel, 1971; Halgren[3]; Leaf, Kayser, Andrews, Adkins, & Leaf, 1968; Lubow & Siebert, 1969; May, Tolman, & Schoenfeldt, 1967; Mellgren & Ost, 1972; Rescorla, 1971; Schnur, 1971; Siegel & Domjan, 1971, Experiment 1), and with some difficulties, and a number of failures, even in humans. The successful demonstrations of latent inhibition in humans are studies by Schnur and Ksir, 1969; Sokolov and Paramanova, 1956; Wolff and Maltzman, 1968; and by Cantor and his associates using children (Bogartz & Witte, 1966; Cantor & Cantor, 1964, 1965, 1966; Cantor & Fenson, 1968; Miller[4]; Witte & Cantor, 1967;

---

[1] The final preparation of this paper was accomplished while the author had a visiting Appointment in Psychology at Bar-Ilan University, Israel. Earlier work on the paper was, in part, supported by National Institute of Health Career Development Award K3 MH-7189, and much of the research by myself and my students by National Institute of Mental Health Grant 5 R01 MH-12307. The author thanks S. E. Newman and P. Schnur for many helpful suggestions.

[2] Requests for reprints should be sent to R. E. Lubow, Department of Psychology, Tel-Aviv University, Ramat-Aviv, Israel.

[3] Halgren, C. R. Latent inhibition: Associative or nonassociative? Unpublished doctoral dissertation, Indiana University, 1971.

[4] Miller, F. D. The effects of differential amounts of stimulus familiarization on discriminative reaction time performance in children. Unpublished doctoral dissertation, University of Iowa, 1966.

TABLE 1

Distribution of Positive and Negative Latent Inhibition Studies Across Species and Tasks

Task	Species						
	Goldfish	Rat, mouse	Rabbit	Goat, sheep	Cat dog	Human child	Human adult
Classical conditioning							
Leg flexion				$+^{1}+^{1}+$			
Pinna or tail response		+	$+^{2}+^{2}$				
Eyelid			$+^{3}+^{3}+^{4}+++-$				$-^{6}-^{6}-+$ ---
Salivary					+		
Fear-bar press suppression		$+^{4}+^{5}+^{5}+^{12}$ ++++					
Fear-lick rate suppression		$+^{7}+^{7}+^{7}+$ $+^{8}+^{8}+^{9}+^{9}+$					
Fear-running speed suppression		$-^{12}$					
Fear-avoidance-escape	+	+++			–		
Gustatory aversion		+++++ ++++					
Galvanic skin response							++
Ivanov-Smolensky							++
Operant-go/no-go discrimination		$+++^{10}+^{10}+^{10}$					
Response speed						$+^{11}+^{11}+++$ ++++++	–

Note. 1–11: The same number next to more than one entry signifies independent experiments reported in the same article. 12: Different dependent measures from the same subjects.

Witte[5]). With the exception of Schnur and Ksir (1969), who used an eyelid-conditioning procedure, all the other human studies depart from traditional conditioning methodology. The failures to find latent inhibition in studies using human adults subjects include Grant, Hake, Riopelle, and Kostlan (1951), Grant, Hake, and Schneider (1948), Perlmuter (1966), Allen,[6] and Schnur,[7] all of which did use the eyelid-conditioning paradigm, and Meyer and Joseph (1968) who used a reaction time procedure. Table 1 shows the distribution of successful and unsuccessful demonstrations of latent inhibition across species and tasks.

Distribution across Testing Situations

Classical conditioning—Pavlovian. Latent inhibition has been demonstrated for the leg flexion response in goats and sheep (Lubow, 1965; Lubow & Moore, 1959); tail movement in the rat (Chacto & Lubow, 1967); pinna response in the rabbit (Lubow, Markman, & Allen, 1968); and the eyelid response in the rabbit (Reiss & Wagner, 1972; Siegel, 1969a, 1969b, 1970; Siegel & Domjan, 1971). Failure to obtain latent inhibition within this classical conditioning procedure is reported by Suboski, DiLollo, and Gormezano (1964) using the nictitating membrane response in rabbits. With adult human subjects using eyelid conditioning, negative findings are reported by Grant et al. (1948), Grant et al. (1951), Perlmuter (1966), Allen (see Footnote 6), and Schnur

---

[5] Witte, K. L. Children's instrumental response speed as related to conditioned stimulus duration during a familiarization experience. Unpublished master's thesis, University of Iowa, 1965.

[6] Allen, J. A. Latent inhibition: Generalization during eyelid conditioning. Unpublished master's thesis, North Carolina State University, 1967.

[7] Schnur, P. Latent inhibition: The effects of nonreinforced preexposure of the CS in differential eyelid conditioning. Unpublished master's thesis, North Carolina State University, 1967.

(see Footnote 7). Only Schnur and Ksir (1969) have reported latent inhibition in humans with eyelid conditioning. While with infrahumans, 12 of 13 classical conditioning experiments have yielded a latent inhibition effect, with adults, 6 of 7 studies have failed to find a latent inhibition effect.

Classical conditioning—conditioned fear as a suppressor. Latent inhibition has been demonstrated in a variety of to-be-suppressed responses: licking rate (Carlton & Vogel, 1967; James, 1971; Leaf et al., 1968; Lubow & Siebert, 1969; Schnur, 1971), bar-press rate (Anderson et al., 1969a; Anderson et al., 1969b; Domjan & Siegel, 1971; May et al., 1967; Rescorla, 1971; Siegel & Domjan, 1971), and running speed (Anderson et al., 1969b). The conditioned suppression technique, as applied to latent inhibition, has been employed in 19 independent experiments reported in 13 articles. Of these only the last cited failed to demonstrate the latent inhibition effect.

Avoidance—escape conditioning. Five studies have been performed with this task. Four of the studies demonstrated the latent inhibition effect (Ackil & Mellgren, 1968; Ackil et al., 1969; Braud, 1971; Grant & Young, 1971) and one study failed to demonstrate it (McDaniel & White, 1966).

Discrimination of go/no-go stimuli. This group of studies extends the latent inhibition phenomenon to an operant situation where the unconditioned stimulus is appetitive. With the exception of one study which employed salivary conditioning with dogs (Konorski & Szwejkowska, 1952), all of the animal studies discussed previously have used classical conditioning with a noxious unconditioned stimulus. The procedure used in the three experiments by Mellgren and Ost (1972) differs from the typical latent inhibition study. During some nonreinforced preexposure periods reinforcement is present but not correlated with any stimulus nor contingent upon any response. In addition both the S^D and S^Δ are successively preexposed. Separate analyses of responding to S^D and to S^Δ indicated that the poorer discrimination performance by the preexposed as compared to the nonpreexposed rats was primarily a function of increased responding to S^Δ by the stimulus preexposure groups. This general trend was apparent in all three experiments. Similar findings are reported by Halgren[8] using rats to lever press for a sucrose solution and by Seybert[9] using a runway brightness discrimination task with running speed as the dependent variable.

Gustatory aversion conditioning. A series of studies concerned with the aversion of a relatively novel food as compared to a familiar food following induced illness also may reflect the effects of latent inhibition. A number of studies find that aversion to flavors after induced toxicosis is reduced by familiarization or preexposure to that flavor (Domjan,[10] Farley, McLaurin, Scarborough, & Rawlings, 1964; Garcia & Koelling, 1967; Maier, Zahorik, & Albin, 1971; McLaurin, Farley, & Scarborough, 1963; Nachman, 1970; Revusky & Bedarf, 1967; Revusky & Garcia, 1970; Wittlin & Brookshire, 1968).

Ivanov–Smolensky conditioning. Conditioning by the method of Ivanov–Smolensky (1933) is an adaptation of the classical Pavlovian procedures to humans. In this technique the conditioned stimulus is paired with a verbal command which serves as the unconditioned stimulus.

There have been only two studies which used this technique in conjunction with nonreinforced preexposure (Siebert[11]; Sokolov & Paramanova, 1956). The latter study, using adolescents and an electromyogram response, as reported by Sokolov (1963), indicates that nonreinforced preexposure results in a severe handicap to subsequent conditioning. The study by Siebert, employing college age subjects and anticipatory thigh muscle movements, although only obtaining low levels of conditioning for all groups, also demonstrated the latent inhibition effect.

---

[8] Halgren, C. R. Latent inhibition: Associative or nonassociative? Unpublished doctoral dissertation, Indiana University, 1971.

[9] Seybert, J. A. Discrimination learning following preexposure to the discriminative stimuli. Unpublished master's thesis, University of Oklahoma, 1971.

[10] Domjan, M. The CS preexposure effect without habituation of attentional responses. Paper presented at the meeting of the Eastern Psychological Association, New York, January, 1971.

[11] Siebert, L. E. The role of the orienting response in latent inhibition using Ivanov–Smolensky conditioning. Unpublished master's thesis, North Carolina University, 1967.

Conditioning of autonomic responses in humans. Sokolov (1963, p. 249) briefly refers to the necessity to reevoke the galvanic skin response (as one member of the orienting response family) after nonreinforced trials before conditioning can occur, and that it takes more trials to evoke the galvanic skin response after nonreinforced preexposure than without nonreinforced preexposure. Wolff and Maltzman (1968) used three groups of undergraduates receiving either 0, 20, or 40 nonreinforced preexposures to word stimuli. The 0-nonreinforced preexposure group reached a peak galvanic skin response conditioned response at the fourth conditioned stimulus, 20 nonreinforced preexposures at the seventh conditioned stimulus, and 40 nonreinforced preexposures at the sixth conditioned stimulus. Surwit[12] presented groups of subjects with either 50 or 100 nonreinforced preexposures to the to-be-conditioned stimulus or to an irrelevant stimulus. The 50- and 100-nonreinforced preexposure groups exhibited fewer galvanic skin response conditioned responses than the groups exposed to the irrelevant stimulus. Amplitude, magnitude, and latency of the galvanic skin response were also analyzed but showed no significant differences between groups.

Response speed. Cantor (1969) has reviewed the stimulus familiarization effect on motor task behavior. His review deals primarily with children and with reaction time measures. Stimulus familiarization effect, "responding less quickly in certain motor tasks when the signal to respond is one with which the subject has been familiarized as opposed to a different one not involved in any prior familiarization operation [Cantor, 1969, p. 144]," has been demonstrated in 11 experiments with children. Five of these studies utilized lever pulling (Cantor & Cantor, 1964; Kubose, 1970; Witte (see Footnote 5); Witte, 1967; Witte & Cantor, 1967), one study used simple reaction time (Bogartz & Witte, 1966), and five used a choice reaction time situation (Bogartz & Witte, 1966, Experiment 2; Cantor & Cantor, 1965, 1966; Cantor & Fenson, 1968; Miller, see Footnote 4). In contrast to these 11 studies with children, all demonstrating the latent inhibition effect, there is one study (Meyer & Joseph, 1968) with female college students which reports no stimulus familiarization effect.

PARAMETERS OF LATENT INHIBITION

Number of Preexposures

The range of *n*-nonreinforced preexposures across studies goes from 1 and 2 (Domjan & Siegel, 1971; Miller, see Footnote 4) to 280 (Suboski, DiLollo, & Gormezano, 1964), 360 (McDaniel & White, 1966), 550 (Siegel, 1970), 1,300 (Siegel, 1969a, 1969b), and 4,176 (May et al., 1967). Most of the other studies use an *n*-nonreinforced preexposure of between 5 and 80, with the typical *n*-nonreinforced preexposure being 20 or 40.

There are 12 studies that specifically investigated the effects of *n*-nonreinforced preexposures. That is, they had at least two levels of *n*-nonreinforced preexposures in addition to the 0-nonreinforced preexposure control condition. Table 2 presents the 12 studies with *n*-nonreinforced preexposure levels for each group in each study. The numbers in the columns refer to the *n*-nonreinforced preexposure for a particular experimental group. A number not underlined indicates that group which exhibited a significant decrement in performance relative to the 0-nonreinforced preexposure group. A number underlined indicates no difference between that group and the 0-nonreinforced preexposure group.

The conclusions from inspecting Table 2 are quite straightforward, yet somewhat startling in that the data from which they are drawn are almost completely consistent between experiments. Of 14 different groups from 6 experiments only 1 (Cantor & Cantor, 1966) shows an inhibitory effect with less than 17 nonreinforced preexposures. On the other hand, within the very narrow range of 16–20 nonreinforced pre-

---

[12] Surwit, R. S. The anticipatory modification of the conditioning of a fear response in humans. Unpublished doctoral dissertation, McGill University, 1972.

TABLE 2

STUDIES INVESTIGATING THE EFFECTS OF NUMBER OF NONREINFORCED PREEXPOSURES ON LATENT INHIBITION

Experiment	Number of nonreinforced preexposures								
	1–5	6–10	11–15	16–20	21–30	31–40	41–50	51–100	101+
Ackil & Mellgren (1968)	5							60	
Cantor & Cantor (1966)	5					35			
Cantor & Fenson (1968)	5	10	12 14	16 18					
Chacto & Lubow (1967)				20		40			
Domjan & Siegel (1971)	2 5	10			25		50		
Grant el al. (1951)	5	10							
Lubow (1965)				20		40			
Lubow et al. (1968)						40		80	
Mellgren & Ost (1972)								75	150 300
Miller (1966)	1	10		20		40			
Siegel (1969a)								100	1300
Wolff & Maltzman (1968)				20		40			

Note. Numbers in the columns refer to *n*-nonreinforced preexposure. Underlined numbers signify failure to obtain LI; numbers not underlined signify success in demonstration of LI.

exposures, suddenly 5 of 6 groups from 5 different experiments show inhibitory effects. This consistency is all the more surprising when one remembers the diversity of the 12 studies in both species and learning paradigm.

Stability of the Latent Inhibition Effect—Temporal Duration

Temporal duration is defined as the time between the last nonreinforced preexposure and the beginning of acquisition. One of the few animal studies which did not obtain a latent inhibition effect used such a 24-hour delay (Suboski et al., 1964). Several studies have followed which specifically investigated the role of temporal delay. Lubow et al. (1968), using rabbits and a conditioned pinna response, with 0- and 24-hour delay, found the delay variable to be significant only in higher order interactions with nonreinforced preexposure, Blocks and Days. Siegel (1970), using rabbits and the conditioned eyelid response, found the usual latent inhibition effect as a function of preexposure but no significant differences as a function of time of delay, 0 and 24 hours. A study by James (1971), using 0-, 30-, and 60-minute time of delay and a conditioned effect response procedure, provides support for Siegel (1970). Several studies used the suppression technique with a 48-hour delay between nonreinforced preexposure and conditioned stimulus-unconditioned stimlus pairing and still report marked latent inhibition effects (e.g. Carlton & Vogel, 1967; Lubow & Siebert, 1969).

Stimulus Specificity of Latent Inhibition

An important factor for explaining the latent inhibition effect is the knowledge of whether nonreinforced preexposure of the to-be-conditioned stimulus results in a general decrement in subsequent performance or whether the inhibitory effects are limited only to that stimulus which has been preexposed. The studies performed in this area each indicate unequivocally that the latent inhibition effect is stimulus specific (Carlton & Vogel, 1967, Experiment 2; Lubow & Moore, 1959; Schnur & Ksir, 1969; Siegel, 1969a).

There is another aspect to stimulus specificity, the place of preexposure in relation to place of testing. A series of studies by Anderson and his associates, using conditioned suppression of bar pressing, were designed to test the effects of place preexposure, conditioning, and testing on the strength of fear conditioning and latent inhibition (Anderson et al., 1968; Anderson et al., 1969a, 1969b). It appears that latent inhibition is increased when the place of preexposure is the same as the place of testing. However, latent inhibition can be achieved when the two places are distinctly different (May et al., 1967).

Grant and Young (1971) demonstrated that preexposure to the test apparatus facilitates subsequent learning of an avoidance response as compared to a group not having prior experience with the apparatus. On the other hand, preexposure to the conditioned stimulus within the apparatus resulted in the same avoidance acquisition scores as for the nonpreexposed-apparatus group. Mellgren and Ost (1972, Experiments 1 & 3), using a go/no-go discrimination, report a similar finding. Since the typical procedure for demonstrating latent inhibition is to preexpose both the *n*-nonreinforced preexposure groups and 0-nonreinforced preexposure group to the apparatus, it may well be that the decremental learning effects from conditioned stimulus preexposure are more pronounced than is currently recognized, being partially masked by the facilitative effects of the apparatus preexposure.

EXPLANATIONS OF LATENT INHIBITION

Habituation of the Orienting Response

The assertion that acquisition of a conditioned response is facilitated by the presence of the orienting reflex is supported in both the Russian literature (Sokolov, 1963; Voronin, Leontiev, Luria, Sokolov, & Vinogradova, 1965) and the Western literature (Maltzman, 1967; Maltzman & Raskin, 1965; Wolff & Maltzman, 1968). If learning is facilitated by the presence of the orienting reflex, then any procedure that reduces the magnitude of the orienting reflex should also decrease the effectiveness of learning. For example, habituation of the galvanic skin response to a list of words prior to conditioning reduces the level of galvanic skin response conditioning performance to a related word (Wolff & Maltzman, 1968).

If latent inhibition is the result of nonreinforced preexposure extinguishing the orienting reflex, any condition which reactivates the orienting reflex should restore normal conditionability. Schnur (1971), in two experiments, using rats and the conditioned emotional response paradigm, tested this deduction. In both cases the hypothesis was not confirmed.

Selective Filters

Schnur (1971) also tested the Mackintosh (1965) derived latent inhibition explanation. The prediction is that nonreinforced preexposure to one element of a later-to-be-compounded stimulus should enhance conditioning to the other element relative to a compound stimulus where both elements are novel. However, even if the position is proven basically correct, it would only be applicable to those few latent inhibition studies that used within-subject designs. Both experiments reported by Schnur (1971) indicate that preexposing one conditioned stimulus element of a compound does not enhance conditioning to the other element as would be predicted by selective attention theory. Sutherland and Mackintosh (1971, p. 68) also evince a puzzlement about how to handle the latent inhibition data.

Competing and Complementing Responses

A different approach to finding an explanation for latent inhibition is to look at overt responses occurring during the nonreinforced preexposure period which are elicited by the to-be-conditioned stimulus and which may either complement or inhibit conditioning performance. Lubow and Moore (1959) suggested that latent inhibition was a result of learning a response to the nonreinforced preexposure stimulus that was incompatible with a later-to-be-conditioned response. However, a test of this hypothesis using leg flexion and extension in sheep and goats failed to confirm the predictions. Likewise, Lubow et al. (1968) showed that latent inhibition was not the result of the habituation of alpha responses during nonreinforced preexposure.

Conditioned Inhibition

Rescorla (1971) defines conditioned inhibition as a "stimulus which comes through learning to specifically interfere with excitation [p. 77]." He suggests dual testing criteria for identifying the presence of a conditioned inhibitor: summation with a known excitor and retardation of the development of a conditioned reflex.

In the Rescorla (1971) study, the first experiment confirmed the earlier findings that preexposure retards subsequent conditioning of suppression. However, the summation test for nonreinforced preexposure effects failed to provide evidence for conditioned inhibition.

Rescorla (1971) also suggested that if conditioned inhibition occurred to the preexposed stimulus, then *n*-nonreinforced preexposure should interfere with any subsequent learning to that stimulus, not only in excitatory conditioning but also in inhibitory conditioning. This was tested by transferring from a latent inhibition procedure to a conditioned inhibition procedure. Evidence for nonreinforced preexposure-produced conditioned inhibition was not found. A related study by Reiss and Wagner (1972) provides additional support for the independence of latent inhibition and conditioned inhibition. These conclusions are buttressed by the results of the three studies demonstrating latent inhibition within the go/no-go discrimination paradigm (Halgren [see Footnote 3]; Mellgren & Ost, 1972; Seybert [see Footnote 9]). In these studies, preexposure of the stimulus which was later to become the stimulus for inhibiting responding resulted in an increased responding to S^{Δ} as compared to a 0-nonreinforced preexposure group. A conditioned inhibition explanation of latent inhibition calls for decreased responding.

CONCLUSIONS

The latent inhibition phenomenon is a robust one, appearing in a variety of species and tasks. A number of variables have been found to have orderly effects on latent inhibition, particularly *n*-nonreinforced preexposure. Nevertheless, a review of the several explanations of latent inhibition indicates that none has strong support. What remains is an explanatory residue which currently goes by the name of stimulus salience or attention decrement. Indeed, there is some support for the idea that nonreinforced preexposure effects a central attentionlike process. Carlton (1969) reports that scopalomine administered before the preexposure session of a conditioned emotional response procedure results in subjects having reliably higher conditioning scores. They exhibit more suppression in the test phase than subjects receiving nonreinforced preexposure without the drug. Similar results have been obtained by Oliverio (1968) with mice. Scopolamine evidently reduces the amount of habituation that can accrue from repeated stimulation. This anticholinergic effect, the attenuation of habituation, or conversely the maintenance of attention, has been well documented for a variety of phenomena other than latent inhibition (Carlton, 1969).

Ackil et al. (1969) have added further support for the central attentional effects of *n*-nonreinforced preexposure. Three groups of rats were given either hippocampal lesions, neocortical lesions, or no lesions. These main groups were subdivided into a 30-nonreinforced preexposure and 0-nonreinforced preexposure condition. The neocortical lesioned and nonlesioned group each demonstrated the usual latent inhibition effect in an avoidance conditioning procedure. However, for the hippocampectomized groups there was no difference in performance between those subjects receiving 0-nonreinforced preexposure and those receiving

30-nonreinforced preexposures. Other studies, (see Douglas, 1967; Kimble, 1968), have demonstrated that the hippocampus plays an important role in the modulation of attention, and, specifically, that hippocampectomy results in the inability to attenuate attention-like responses to repeated stimulation.

Since the effects of these centrally operating drugs and lesions are widely regarded as resulting in a failure to produce attention decrements and the same drugs and lesions also prohibit the development of latent inhibition, it is suggested that during nonreinforced preexposure, for the normal, nondrugged, nonhippocampectomized subject, there is a centrally mediated decrement of attention to the preexposed stimulus, and that this decrement is responsible for the subsequent latent inhibition effect.

The explanations of latent inhibition that were examined emphasized *either* a learning or an attentional mechanism. The learning approach sought to discover relatively gross motor responses associated with nonreinforced preexposure such as leg flexor and extensor antagonism, or to look for conditioned inhibitory effects where the stimulus acquires general inhibitory properties which effect the organism's responding to not only that stimulus but to other stimuli with which it might later be compounded. Attentional explanations of latent inhibition simply conferred a degree of reduced stimulus saliency as a result of nonreinforced preexposure. In general, this type of explanation is congruent with the data from the stimulus compounding studies of Reiss and Wagner (1972), Rescorla (1971), and Schnur (1971), and meets the requirements of stimulus specificity. However, it does not suggest a mechanism by which nonreinforced preexposure reduces salience. Furthermore, it does not attempt to deal with the facts of generalization of latent inhibition (Siegel, 1969b), and its relative stability over time (e.g., Lubow & Siebert, 1969; Mellgren & Ost, 1972) but susceptibility to counterconditioning-like procedures (Carlton & Vogel, 1967; James, 1971). These latter considerations point to the need for a model that combines attentional and learning constructs.

REFERENCES

Ackil, J. E., & Mellgren, R. L. Stimulus preexposure and instrumental learning. *Psychonomic Science*, 1968, **11**, 339.

Ackil, J. E., Mellgren, R. L., Halgren, C., & Frommer, G. P. Effects of CS preexposure on avoidance learning in rats with hippocampal lesions. *Journal of Comparative and Physiological Psychology*, 1969, **69**, 739–747.

Anderson, D. C., Merril, H. K., Dexter, W., & Alleman, H. Contextual effects in emotional learning. *Proceedings of the American Psychological Association*, 1968, **3**, 147–148.

Anderson, D. C., O'Farrell, T., Formica, R., & Caponigri, V. Preconditioning CS exposure: Variation in place of conditioning and presentation. *Psychonomic Science*, 1969, **15**, 54–55. (a)

Anderson, D. C., Wolf, D., & Sullivan, P. Preconditioning exposure to the CS: Variation in place of testing. *Psychonomic Science*, 1969, **14**, 233–235. (b)

Bogartz, R. S., & Witte, K. L. On the locus of the stimulus familiarization effect in young children. *Journal of Experimental Child Psychology*, 1966, **4**, 317–331.

Braud, W. G. Effectiveness of "neutral," habituated, shock-related and food-related stimuli as CSs for avoidance learning in goldfish. *Conditional Reflex*, 1971, **6**, 153–156.

Cantor, G. N. Stimulus familiarization effect and the change effect in children's motor task behavior. *Psychological Bulletin*, 1969, **71**, 144–160.

Cantor, G. N., & Cantor, J. H. Effects of conditioned stimulus familiarization on instrumental learning in children. *Journal of Experimental Child Psychology*, 1964, **1**, 71–78.

Cantor, G. N., & Cantor, J. H. Discriminative reaction time performance in pre-school children as related to stimulus familiarization. *Journal of Experimental Child Psychology*, 1965, **2**, 1–9.

Cantor, G. N., & Cantor, J. H. Discriminative reaction time in children as related to amount of stimulus familiarization. *Journal of Experimental Child Psychology*, 1966, **4**, 150–157.

Cantor, G. N., & Fenson, L. New data regarding the relationship of amount of familiarization to the stimulus familiarization effect (SFE). *Journal of Experimental Child Psychology*, 1968, **6**, 167–173.

Carlton, P. L. Brain-aceytlcholine and inhibition. In J. T. Tapp (Ed.), *Reinforcement and behavior*. New York: Academic Press, 1969.

Carlton, P. L., & Vogel, J. R. Habituation and conditioning. *Journal of Comparative and Physiological Psychology*, 1967, **63**, 348–351.

Chacto, C., & Lubow, R. E. Classical conditioning and latent inhibition in the white rat. *Psychonomic Science*, 1967, **9**, 135–136.

Domjan, M., & Siegel, S. Conditioned suppression following CS preexposure. *Psychonomic Science*, 1971, **25**, 11–12.

Douglas, R. J. The hippocampus and behavior. *Psychological Bulletin,* 1967, **67**, 416–442.

Farley, J. A., McLaurin, W. A., Scarborough, B. B., & Rawlings, T. D. Pre-irradiation sacharine habituation. A factor in avoidance behavior. *Psychological Reports,* 1964, **14**, 491–496.

Garcia, J., & Koelling, R. A. A comparison of aversions induced by X-rays, toxins and drugs in the rat. *Radiation Research,* 1967, Suppl. 7, 439–450.

Grant, D. A., Hake, H. W., Riopelle, A. J., & Kostlan, A. Effects of repeated pre-testing with conditioned stimulus upon extinction of the conditioned eyelid response to light. *American Journal of Psychology,* 1951, **54**, 247–251.

Grant, D. A., Hake, H. W., & Schneider, D. E. Effects of pre-testing with the conditioned stimulus upon extinction of the conditioned eyelid response. *American Journal of Psychology,* 1948, **61**, 243–246.

Grant, M., & Young, D. The effect of preexposure to a learning apparatus. *Behavior Research Methods and Instrumentation,* 1971, **3**, 251–252.

Ivanov-Smolensky, A. G. *Methods of investigation of conditioned reflexes in man.* Moscow: State Medical Publishing House, 1933.

James, J. P. Latent inhibition and the preconditioning interval. *Psychonomic Science,* 1971, **24**, 97–98.

Kimble, D. P. Hippocampus and internal inhibition. *Psychological Bulletin,* 1968, **70**, 285–295.

Konorski, J., & Szwejkowska, G. Chronic extinction and restoration of conditioned reflexes. IV. The dependence of the course of extinction and restoration of conditioned reflexes on the history of the conditioned stimulus (the principle of primacy of first training). *Acta Biologiae Experimentalis,* 1952, **16**, 95–113.

Kubose, S. K. The stimulus familiarization effect as a function of stimulus similarity. *Psychonomic Science,* 1970, **18**, 367–368.

Leaf, R. C., Kayser, R. J., Andrews, J. S., Jr., Adkins, J. W., & Leaf, S. R. P. Block of fear conditioning induced by habituation or extinction. *Psychonomic Science,* 1968, **10**, 189–190.

Lubow, R. E. Latent inhibition: Effects of frequency of non-reinforced preexposure to the CS. *Journal of Comparative and Physiological Psychology,* 1965, **60**, 454–455.

Lubow, R. E., Markham, R. E., & Allen, J. Latent inhibition and classical conditioning of the rabbit pinna response. *Journal of Comparative and Physiological Psychology,* 1968, **66**, 688–694.

Lubow, R. E., & Moore, A. U. Latent inhibition: The effect of non-reinforced preexposure to the conditional stimulus. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 416–419.

Lubow, R. E., & Siebert, L. Latent inhibition within the CER paradigm. *Journal of Comparative and Physiological Psychology,* 1969, **68**, 136–138.

Mackintosh, J. J. Selective attention in animal discrimination learning. *Psychological Bulletin,* 1965, **64**, 124–150.

Maier, S. F., Zahorik, D. M., & Albin, R. W. Relative novelty of solid and liquid diet during thiamine deficiency determines development of thiamine-specific hunger. *Journal of Comparative and Physiological Psychology,* 1971, **74**, 254–262.

Maltzman, I. Individual differences in "attention": The orienting reflex. In R. M. Gagne (Ed.), *Learning and individual differences.* Columbus, Ohio: Merrill, 1967.

Maltzman, I., & Raskin, D. C. Effects of individual differences in the orienting reflexes on conditioning and complex processes. *Journal of Experimental Research in Personality,* 1965, **1**, 1–16.

May, R. B., Tolman, C. W., & Schoenfeldt, M. G. Effects of pretraining exposure to the CS on conditioned suppression. *Psychonomic Science,* 1967, **9**, 61–62.

McDaniel, J. W., & White, R. K. A factorial study of the stimulus conditions of habituation. *Perceptual Motor Skills,* 1966, **23**, 259–270.

McLaurin, W. A., Farley, J. A., & Scarborough, B. B. Inhibitory effects of pre-irradiation sacharine habituation on conditioned avoidance behavior. *Radiation Research,* 1963, **18**, 473–478.

Mellgren, R. L., & Ost, J. W. P. Discriminative stimulus preexposure and the learning of an operant discrimination. *Journal of Comparative and Physiological Psychology,* 1972, **77**, 179–187.

Meyer, W. J., & Joseph, L. J. Response speed as related to CS prefamiliarization and GSR responsivity. *Journal of Experimental Psychology,* 1968, **78**, 375–381.

Nachman, M. Learned taste and temperature aversions due to lithium chloride sickness after temporal delays. *Journal of Comparative and Physiological Psychology,* 1970, **73**, 22–30.

Oliverio, A. Effects of scopolamine on avoidance conditioning and habituation of mice. *Psychopharmacologia,* 1968, **12**, 214–216.

Perlmuter, L. C. Effect of CS manipulation on the conditioned eyelid response: Compounding, generalization, in the inter-CS-interval, and preexposure. *Psychonomic Science Monographs,* 1966, **1**, 271–286.

Reiss, S., & Wagner, A. R. CS habituation produces a "latent inhibition effect" but no active "conditioned inhibition." *Learning and Motivation,* 1972, **3**, 237–245.

Rescorla, R. A. Summation and retardation tests of latent inhibition. *Journal of Comparative and Physiological Psychology,* 1971, **75**, 77–81.

Revusky, S., & Garcia, J. Learned associations over long delays. In G. Bower (Ed.), *The psychology of learning and motivation. Vol. 4.* New York: Academic Press, 1970.

REVUSKY, S. H., & BEDARF, E. W. Association of illness with prior ingestion of novel foods. *Science*, 1967, **155**, 219–220.

SCHNUR, P. Selective attention: Effect of element preexposure on compound conditioning in rats. *Journal of Comparative and Physiological Psychology*, 1971, **76**, 123–130.

SCHNUR, P., & KSIR, C. J. Latent inhibition in human eyelid conditioning. *Journal of Experimental Psychology*, 1969, **80**, 388–389.

SIEGEL, S. Effects of CS habituation on eyelid conditioning. *Journal of Comparative and Physiological Psychology*, 1969, **68**, 245–248. (a)

SIEGEL, S. Generalization of latent inhibition. *Journal of Comparative and Physiological Psychology*, 1969, **69**, 157–159. (b)

SIEGEL, S. Retention of latent inhibition. *Psychonomic Science*, 1970, **20**, 161–162.

SIEGEL, S. Latent inhibition and eyelid conditioning. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II*. New York: Appleton-Century-Crofts, 1971.

SIEGEL, S., & DOMJAN, M. Backward conditioning as an inhibitory procedure. *Learning and motivation*, 1971, **2**, 1–11.

SOKOLOV, E. N. *Perception and the conditioned reflex*. New York: Pergamon Press, 1963.

SOKOLOV, E. N., & PARAMANOVA, N. P. Concerning the role of the orientation reflex in the formation of motor conditioned reactions in man. *Journal of Higher Nervous Activity*, 1956, **6**, 702–709. Cited by E. N. Sokolov, *Perception and the conditioned reflex*. New York: Pergamon Press, 1963.

SUBOSKI, M. D., DILOLLO, V., & GORMEZANO, I. Effects of unpaired preacquisition exposure of CS and UCS on classical conditioning of the nictitating membrane response of the albino rabbit. *Psychological Reports*, 1964, **15**, 571–576.

SUTHERLAND, N. S., & MACKINTOSH, N. J. *Mechanisms of animal discrimination learning*. New York: Academic Press, 1971.

VORONIN, L. G., LEONTIEV, A. N., LURIA, A. R., SOKOLOV, E. N., & VINOGRADOVA, O. S. (Eds.), *Orienting reflex and exploratory behavior*. Washington, D. C.: American Institute of Biological Sciences, 1965.

WITTE, K. L. Children's response speeds to familiarized stimuli. *Psychonomic Science*, 1967, **7**, 153–154.

WITTE, K. L., & CANTOR, G. N. Children's response speeds to the offset of novel and familiar stimuli. *Journal of Experimental Child Psychology*, 1967, **5**, 372–380.

WITTLIN, W. A., & BROOKSHIRE, K. H. Apomorphine induced conditioned aversion to a novel food. *Psychonomic Science*, 1968, **12**, 217–218.

WOLFF, C., & MALTZMAN, I. Conditioned orienting reflex and amount of preconditioning habituation. *Proceedings of the American Psychological Association*, 1968, **3**, 129–130.

(Received July 25, 1972)

Psychological Bulletin
1973, Vol. 79, No. 6, 408-416

THEORIES OF DREAM FUNCTION

JANET DALLETT[1]

University of California, Los Angeles, and C. G. Jung Clinic, Los Angeles

Some contemporary and psychoanalytic theories of dream function are reviewed. Particular attention is given to Jung's model of personality and theory of dream function, a dynamic, open-system approach that stands in contrast to Freud's mechanistic, drive-reduction model. Contemporary theories tend to focus on the function of environmental mastery, viewed from one of three perspectives: (*a*) problem solving, (*b*) information processing, or (*c*) ego consolidation. Only a few have gone beyond environmental control to consider creative functions of the dream. Most contemporary theories are at least partially supported by data, and many are not mutually exclusive, dealing with different processes or proposing similar processes in different-sounding languages. Some dream theories do differ from others in underlying model, in scope, or in the degree to which nonrational processes are admissable as data. Jung's approach has much to add to contemporary dream theory, particularly in making room for creative and nonrational processes, as well as in the specific proposition that dreams function to balance and complete waking consciousness.

Many answers have been suggested to the question of what function dreams serve in man's psychology. In this century, the most influential theory has been Freud's and his views have given laboratory psychology a fruitful source of hypotheses and debate. The mainstream of American scientific thought in this area has grown directly from laboratory research but is richly sprinkled with Freudian theory together with a Darwinian emphasis on adaptation as environmental mastery. Jung's approach to dream function, very different from Freud's and perhaps more compatible with current research findings, has remained outside the mainstream and is virtually unknown to American psychology. This paper aims to correct the omission, while providing an overview of psychoanalytic and contemporary thought about the function of dreams.

Freud (1953) believed that dreams serve a dual, compromise function. According to his theory, unconscious, instinctual drive energy pushes for discharge, moving toward the expression of a consciously unacceptable impulse. The reduction in conscious restraints characteristic of sleep permits a symbolic, disguised dream expression of the repressed wish. The overt (manifest) content of the dream represents a compromise between the instinctual forces (latent content) striving for expression, on the one hand, and the repressive forces of consciousness, on the other. Freud assumed that the energy pushing for action would awaken the sleeper were it not for the dream which, through symbolic discharge, permits a return to sleep. Thus, the dream is seen as serving the biological function of preserving sleep, as well as the psychological function of discharging an unacceptable wish that might otherwise burst destructively into waking life.

Various aspects of Freud's dream theory have been reviewed from the point of view of contemporary dream research (Breger, 1967; Fisher, 1965; Foulkes, 1964; Hollander, 1962; Jones, 1968, 1970; Trosman, 1963), and of changes in the philosophy of science.[2] While conclusions vary, it is generally agreed that at least some aspects of the theory must be discarded or revised. With respect to dream function in particular, the sleep preservation view is untenable and the underlying model upon which the wish-fulfillment theory rests requires extensive revision.

It must be emphasized that the psychology of Freud's theoretical model rests on his distinction between manifest and latent dream content, that is, the assumption that meaning is concealed rather than revealed by the dream. Granting the puzzling, chaotic, and foreign flavor of much dream content, it is more straightforward to assume that the dream is an expression lacking the discriminative order imposed by conscious ("secondary process") thought, and which must therefore be ordered and decoded if it is to be understood within the framework of normal adult thought. Whether dreams disguise, reveal, or are meaningless must remain a question of faith, but most theorists, including many whose work builds upon Freud's (e.g., Adler, Eriksen, Jung, Maeder, and Silberer) incline toward the view that they reveal.

---

[1] Requests for reprints should be sent to Janet O. Dallett, 929 Gretna Green Way, Los Angeles, California 90049. The author is indebted to Louis Breger, Seymour Feshbach, Joseph Henderson, and Yechezkel Kluger for their critical reading and helpful comments contributing to the final form of this paper.

[2] P. Castle. The significance of thought and action in dreaming: A theoretical analysis. Unpublished doctoral dissertation, Brandeis University, 1970.

Despite the many problems inherent in his theoretical formulation of dream function, Freud's far-reaching work has provided a fruitful basis for many of the contemporary theories discussed below. Psychologists steeped in the Freudian tradition have generated theories compatible with those arising from the laboratory alone, generally moving away from wish-fulfillment and drive-discharge functions toward an emphasis on adaptation and ego synthesis. Forerunners of this trend are to be found among Freud's original psychoanalytic group and, indeed, can be discerned in Freud's own brilliant interpretative work which, as Jones (1970) has emphasized, often ranges far from the letter of his theory.

A large group of investigators approach the dream as essentially continuous with waking thought. While specifics of the theories vary, each in its own way attributes to the dream the function of facilitating some aspect of the dreamer's mastery over his environment. Environmental mastery is usually approached from one of three points of view: (*a*) problem solving, (*b*) information processing, or (*c*) ego consolidation.

The problem-solving view of dream function was already important to some of Freud's early followers. Maeder (1913), while he believed with Freud that dreams have a drive-discharge function, also saw them as tentative, practice attempts to solve problems. Adler (1936) saw the dream as an expression of the dreamer's characteristic approach to life, his unresolved reality problems and, with Maeder, as an attempt to meet and solve these problems. In addition, Adler believed that dreams arouse emotions that push the dreamer toward problem solution. However, he felt that the *solutions* embodied in dreams are self-deceptive, constituting a flight from realistic coping with life.

Ullman (1959, 1960, 1962) has built upon Adler's theory but considers that dream content itself constitutes an adaptive contribution to problem solving, rather than a flight from reality. He feels that the dream expresses symbolically the connections between a present problem and relevant aspects of past experience; it permits a fuller assessment and thereby better solution to the problem than is available to waking consciousness alone. Ullman also looks at the adaptive value of the dream for species survival, assuming that the apparent lightening of sleep associated with the dream state originally evolved out of the necessity for primitive man to arouse periodically in order to survive physical danger. Dream images form an assessment of currently disturbing life situations and, depending on the outcome, the dream may either facilitate full arousal or promote a return to deeper sleep.

Klein (1967) also maintains a close relation between daytime thoughts and the dreams of the night, seeing dreams as one mode of carrying on a series of ideas that may have been set in motion during the day. There is a certain amount of interchangeability among thoughts, fantasies, dreams, and actions in a "series of ideomotor events" that, once set in motion, pushes for completion in and of itself, not in the service of drive reduction. Building on this model, Fiss (1969) conceptualizes the need to dream as a need to complete unfinished tasks. Unresolved (incomplete) cognitive-emotional tasks constitute problems which continue toward completion during the night.

Pearlman (1970), while he, too, believes that dreams play a role in solving problems, distinguishes between their role and that of waking thought. He points out that intellectual insight is different from the emotional kind of "knowing" that makes an insight a part of oneself and proposes that dreams function specifically to facilitate the emotional assimilation that gives adaptive reality to a problem solution.

Breger (1969) emphasizes the problem-solving aspect of dreams for children in particular, arguing for a developmental function facilitating the child's efforts at mastery. He follows Griffiths (1935) in pointing out that fantasy is the natural mode of thought in early childhood and the primary mode through which children work through emotional problems. Like Piaget, Breger views dreaming as an assimilative form of thought which allows free "experimentation" with problem situations and out of which emerge "fantasy programs" for creative new solutions.

In another paper, Breger (1967) applies the problem-solving approach to an information-processing model and refers it to the solution of affective-personal problems in general, over and above those of child development. He hypothesizes that in dreams, current, affectively arousing problem situations are compared and "tested for fit" with various past "programs" that have served more or less satisfactorily to resolve earlier conflicts. Although couched in different language, this is similar to Ullman's approach and also to the views of French (1952) and French and Fromm (1964), who see dreams as attempted solutions of "focal conflicts." Likewise Greenberg (1970) suggests that in dreams, past and present experiences of a similar emotional nature are brought together and "filed," allowing past experience to facilitate new adaptation and new experience to alter old modes of coping. Breger (1967) points out that the dream state has a number of advantages over waking consciousness for dealing with emotional material: Stored information is more readily available; associational processes are more fluid; the criterion of social acceptability is at a minimum; and a greater variety of means for manipulating symbols is available. The overall effect is a "creative opening up of the memory systems [p. 25]."

Shapiro (1967, 1968) proposes a somewhat different information-processing model. Sleep and dreaming are seen as parts of a total organismic process whereby the nervous system comes increasingly to approximate a "map" of external reality as a result of experience, with corresponding changes in attitudes, feelings, etc. There are two modes of reality testing: the "computing," which corresponds roughly to secondary-process thinking, and the "comparing," which approximates primary-process thinking. Both modes of thought operate during sleep and waking but the former is more dominant during waking, the latter during sleep when new data are compared for relevance with past data and future expectations. A feedback system normally produces alternation between the two modes of thought, both of which are necessary for complete processing of incoming experience. The noxious effects of sleep and/or dream deprivation result from the accumulation of unprocessed data.

Hawkins (1966), in a still different variation of the information-processing theme, conceives the waking state as one of constant input, the dream state as one of sorting and storage required after a certain amount of input has occurred. At the physiological level the sorting-storage process can be seen as central nervous system reorganization. Translated into psychoanalytic language, it is a process of ego recathexis but not drive discharge.

In fact, a number of the theorists who propose an environmental-mastery function approach dreams via the concept of ego consolidation. Eriksen (1954) has spoken of the "identity-preserving function of the dream," a view that Jones (1962) has elaborated. According to Jones, the dream advances ego synthesis through its integration of new experiences from waking life into a process whereby the dreamer's "nuclear crises" are resolved. Similarly, Castle (see Footnote 2), who follows Piaget in understanding the dream as egocentric and assimilative thought, suggests that dreams maintain ego continuity by relating new material to the dreamer's past history, thereby allowing the integration of unacknowledged material which would otherwise remain unassimilated. Following the same line of thought, Witkin (1969) hypothesizes that in the course of a dream, the dreamer's position with respect to the central problem is likely to improve, putting him in a position of relatively more mastery than at the beginning of the dream. Witkin emphasizes the dimension of psychological differentiation as a possible equivalent of ego development and cites unpublished work by Eagle showing that the dreamer is more often represented as an active participant (implying relatively more ego development) in the dreams of more differentiated persons.

Lerner (1967) points out that, while differentiation is important, very little attention has been given to the related question of the "reintegration" of experience. Her view, an outgrowth of Rorschach's work, is that dreams contribute to ego development by functioning to reintegrate the body into fantasies which the dreamer cannot act out in waking life. Body image is seen as the basis for ego and, following Rorschach, Lerner proposes that kinesthetic fantasy is required to maintain coherence of body image, hence of personality organization. Immobility is required for the production of kinesthetic fantasy, a condition that is met by the dream state but rarely by other natural conditions.

Some of the mastery-oriented views of dream function concern themselves also with its creative aspects insofar as creativity is a tool in the practical business of gaining control over the environment. Only a few contemporary theorists have approached the contribution of dreams to the creative process as a value in and of itself, adding something new to human life beyond the utilitarian function of survival and environmental control.

Krippner and Hughes (1970) view information processing secondarily as an *early* function of dreams, which in later life continues to employ a mode of thought that is characteristic of childhood mental processes. This mode rarely appears in adult thinking but is a vital aspect of creative thought. These authors cite studies showing that creative people tend to remember more dreams than noncreative people and to have more exotic dreams; they also mention a number of historical personalities whose discoveries, inventions, and/or artistic products were mediated by dreams.

Piaget (1951) is noteworthy in his view that dreams, along with play, constitute a creative process of construction, an essential aspect of the equilibrative mechanism through which human cognitive development occurs. He recognizes the subjective, affective concerns that are maximally operative in dreams as no less important than environmental reality in the interaction between organism and environment. Like play, dreams may occur in a dynamically compensatory relation to environmental reality. Ironically, the views of Piaget and of Jung, who have been sharply critical of one another, are in many respects more similar than the thought of either of these two Swiss psychologists is to American or to Freudian psychology.

Like Freud's, Jung's theory of dream function is an integral part of a larger view of personality, although it differs quite radically from Freud's. So far as specific functions of the dream are concerned, Jung's view incorporates many of those described above. As Ellenberger (1970) has pointed out,

> ...Jung maintains that the functions of dreams are manifold. They can express fears as well as wishes; they can give a mirror picture of the dreamer's actual situation; there are prospective dreams; others are creative, warning, or parapsychological [p. 716].

Jung does, however, propose an overriding principle of dream function within which the specifics operate.

With the rare exceptions noted above, contemporary theories of dream function are characterized by an exclusive concern with adaptation to the world of concrete, outer reality. Jung gives equal emphasis to the world of subjective, inner experience. Further, unlike both Freud and contemporary theorists, Jung retains a deep concern for the irrational side of man's nature. His attitude is that the irrational *is*, whether we like it or not, and any approach to psychology that tries to exclude it is unable to explain, or even to describe, a great deal of human behavior. Further, through reductionism the richness of human life and values becomes lost.

Jung observed, systematized, and named the phenomena he observed but did not consider himself to be a theorist. Working from observation rather than theory, he felt no necessity to fit later observations into earlier formulations, nor to document his inconsistencies. He said (1953), "I ... prefer the precious gift of doubt, for the reason that it does not violate the virginity of things beyond our ken [p. 8]." Such fluidity has obvious advantages, together with the disadvantage (among others) that it is difficult to become familiar enough with his thinking not to misrepresent it. Nevertheless, an attempt is made here to set down the broad outlines of a model of personality functioning, including dream function, that emerges from and underlies his observations. Because Jung's view is little known, it is described in some detail.

For Jung a construct called "the unconscious" is central. Jung's view of the unconcious is different from Freud's and the two must be distinguished. It is difficult to define the unconscious in a precise way. As Jung said in a filmed interview, "The trouble with the unconscious is that it *is* unconscious!" Looking at the unconscious is a little like looking at the back of one's own neck; it cannot be done directly, but only through an arrangement of mirrors, leaving doubt as to its "true" nature. It is particularly difficult to define in the context of a psychology that has only recently begun to emerge from an empty-organism view to which the notion of consciousness is hardly acceptable, without which the contrasting unconscious cannot be defined. That is, the unconscious comprises "the totality of all psychic phenomena that lack the quality of consciousness [Jung, 1960a, p. 133]."

Nevertheless, as a beginning we can say that the unconscious includes all those aspects of the total personality that are not under conscious control. This includes not only what was once conscious and has been repressed, but also innate potentialities that have not yet become conscious. Obviously, at a superficial level what is unconscious varies from moment to moment depending on fluctuations in attention.

Unconscious contents include, for example, forgotten or repressed memories, psychological representations of bodily functions of which we are not aware, and innate and acquired patterns of behavior. The primary mirrors in which these contents can be observed are *affects* and *images*. As the spontaneous activity perhaps furthest from actual or potential conscious control, dream images are one of the best mirrors of the unconscious. As von Franz[3] puts it,

> Our concept of the unconscious is a borderline concept, for we simply describe by this term a real though invisible fact, for ... the reality of the unconscious is a concluded reality, not demonstrable directly. You can say that it is what produces dreams, that it is that "X" which arranges the meaningful order of the dream. With a certain technique of interpretation we can bring out of the dream a meaningful connection, so we conclude that there must be something which produces the dream. We have ... a certain amount of fact and extract therefrom a concept which we call the unconscious [p. 2–2].

Although in their manifestations they are usually mixed together, Jung classifies unconscious contents as of two sorts: personal and collective. The personal unconscious consists partly of the Freudian unconscious, namely, all those contents that have been repressed because they are too painful or incompatible with conscious views to recognize. In addition, there are neutral contents and, more important, memories and parts of the personality which may be acceptable in themselves but have been repressed because they are not in accord with the image of himself that the individual has developed.

The collective unconscious constitutes all those aspects of the psyche that are basic to the human condition, those psychological characteristics that belong to us by virtue of our biological structure. It is part of being human that most of us are born with two legs, two arms, a cerebral cortex, and the capacity to walk upright. These and all other aspects of our biology have certain psychological consequences and lead to typical modes of functioning and patterns of behavior that are reflected in the affects and images of the collective unconscious. Underlying the images are nuclei of energy which give impulse and direction to thought, perception, and at the instinctual level, behavior. These energic nuclei are called archetypes, which are

> the necessary *a priori* determinants of all psychic processes. Just as his instincts compel man to a specifically human mode of existence, so the archetypes force his way of perception and apprehension into specifically human patterns [Jung, 1960a, p. 133].

The concept of the archetype has given rise to much misunderstanding, partly because Jung's formulation changed in the course of his life and partly because Jungians tend to use the term loosely to refer to images. In fact, an image is never the archetype itself, but simply one of the mirrors from which the construct has been inferred. The archetypal image constitutes the specific form in which an archetype may express itself and is not invariant, for it is almost inevitably colored by both cultural and personal determinants. The archetype bears no relation to the notion that acquired ideas can be inherited. The archetype is no more mysterious than the fact that the brain is programmed to work in particular ways and not in others, with particular psychological consequences. For example, one of the most fundamental attributes of the psyche is the tendency to perceive the world in dualities: this and not that, black and white, male and female, high and low, I and not I, etc. It is difficult to conceive of the development of the discriminative, organized thought characteristic of human consciousness without this fundamental tendency as a starting point.

---

[3] M.-L. von Franz. *The dreams and visions of St. Niklaus von der Flüe.* Zürich: C. G. Jung Institute, 1957. (Mimeo)

It is notable that Breger (1967), searching for a model more compatible with known biological reality than is Freud's theory of psychic energy, has arrived at the necessity for a concept which very closely approximates that of the archetype:

> The difficulties stemming from the psychic-energy theory may be in large part avoided by assuming that what is biologically basic ("instinctual" in Freud's terms) is the *direction given to thought and action by the structure or organization of the nervous system.* The ethologists . . . take this approach when they posit structural mechanisms underlying the "fixed action patterns" that comprise the "instinctual behavior" of certain birds and fish. Such structures have evolved, through the process of natural selection, because of the adaptive advantages they confer on the species possessing them. . . . These guiding structures are embodied in the brain and nervous system and have evolved to their present state of complexity over long periods of time, just as have the other biological systems of the body. . . . It is worth stressing that even in the case of the relatively fixed action patterns, it is typically goals or environmental achievements, and not rigid sequences of motor acts, that are built in. . . . These sorts of guiding structures . . . provide the *direction* of ongoing activities that is motivation [pp. 11–14].

As is evident from this brief overview, Jung's view of the unconscious is far from Freud's seething cauldron. While recognizing the dark, negative aspects, Jung sees the unconscious in addition as the often untapped source of positive, adaptive, and creative potential; of man's highest possibilities as well as his lowest.

In Jung's view, the personality can be represented, as dreams sometimes represent it, as a large circle within which consciousness is contained as a very small island. The ego, the "complex of identity" (see Whitmont, 1969), is the center of that tiny island of consciousness, but not of the whole personality.

It is not unusual for the ego to think that consciousness comprises the total personality and that the vast area outside consciousness is either nonexistent or insignificant. The egocentric view either ignores manifestations of the unconscious, reduces them to the point of view of the ego, or projects them outward, explaining them as forces acting from the outside. Jung sees this as a drastically limited and limiting approach that reduces man to an infinitesimal portion of his potentiality.

Jung observed that the content of individual dreams appears to bear a relation best described as *compensatory* to the attitude of ego-consciousness. To understand what he meant by compensation, one must bear in mind the picture of the total personality. The dream presents a point of view from outside consciousness but from within the whole personality. In simple instances the dream content may be directly complementary to the ego's attitude; for example, if the ego says "I am white," the dream may imply that the ego is in fact black. Usually compensation is more subtle, however, and a general statement of the principle is that the dream confronts the ego with what is now most necessary to bring its attitude into accord with the reality of the whole personality, to restore intrapsychic balance. As Jung (1960b) puts it, compensation implies "balancing and comparing different data or points of view so as to produce an adjustment or a rectification, and . . . a self-regulation of the psychic system [pp. 287*f*]." So long as the aims of the ego remain in relative harmony with the whole, the compensatory process appears to proceed automatically, without conscious attention from the ego. However, if the ego persistently behaves in ways that do violence to the whole, the compensatory activity of the unconscious becomes extreme and may result in neurotic symptoms, psychogenic illness, moods, outbursts of affect, accidents, nightmares, etc. All such symptoms can be understood as compensatory responses to an ego attitude that has become too one-sided. When they arise, serious attention to dreams can provide clues to the inner or outer changes required to restore more balanced functioning.

Jung did not believe that the dream "tries" to conceal anything. Rather, he saw the dream as the best possible symbolic (usually not literal) expression of a situation. The language of unconscious symbols is different from that of rational, verbal, or linguistic symbols, however, and a good deal of work is required to understand a dream in its own terms. Jung (1966a) says of this:

> The true symbol differs essentially from [the sign or symptom], and should be understood as an expression of an intuitive idea that cannot yet be formulated in any other or better way. When Plato, for instance, puts the whole problem of the theory of knowledge in his parable of the cave, or when Christ expresses the idea of the Kingdom of Heaven in parables, these are genuine and true symbols, that is, attempts to express something for which no verbal concept yet exists. If we were to interpret Plato's metaphor in Freudian terms we would naturally arrive at the uterus, and would have proved that even a mind like Plato's was still stuck on a primitive level of infantile sexuality. But we would have completely overlooked what Plato actually created out of the primitive determinants of his philosophical ideas; we would have missed the essential point...[p. 70].

Dreams can, and sometimes do, refer to outer situations and people, and they are often understood as such on what Jung calls the objective level. Often, even when a dream contains people and situations from the outer world, it is more meaningful to understand it in a completely inner way, on the so-called subjective level. Taken subjectively, everything and everyone in the dream embodies an aspect of the larger personality. For example, a very unsociable person may dream about a life-of-the-party friend who embodies his own unlived sociability. Whether it is viewed objectively or subjectively, every dream must be seen in the context of the life situation and conscious attitudes of the dreamer as the principle of compensation implies. It is impossible to be certain of the meaning of a dream out of context.

According to Jung, then, individual dreams function in complex ways to compensate ego-consciousness. We cannot say "in order to compensate"; compensation is simply a description of what happens. "A dream never says 'you ought' or 'this is the truth.' It presents an image in much the same way as nature allows a plant to grow, and it is up to us to draw conclusions [Jung, 1966b, p. 104]."

When dreams are observed and "related to" (valued, with an attempt to understand them) over a period of time, several things happen. If the ego-attitude has been extremely one-sided, it may move closer to the center of the whole, in which case neurotic symptoms generally disappear. The conscious personality also broadens, so to speak, as more and more of the inner possibilities are accepted and lived. And one begins to notice, longitudinally, a second function of dreams. The specific compensations of each night begin to reveal a complex pattern that appears *as if* all the conflicting and diverse potentials of the total personality "wanted" to be realized in the life of the individual in what Jung calls the process of individuation. It is *as if* the point–counterpoint between ego and unconscious were together moving toward some ideal goal of the fulfillment of the whole man. Thus Jung infers a developmental function of dreams which moves, through the whole of life, toward the realization of the total personality. Dement (1966) has said of rapid eye movement (REM) sleep,

> never before in the history of biological research has so much been known about something from a descriptive point of view, with so little known at the same time about its function. Yet to doubt that the REM state has a major and vital biological destiny to fulfill is to fly in the face of well-established evolutionary principles [p. 136].

Likewise, given the universality of dreaming, there can be little doubt that dreams themselves play some vital role in man's psychic economy, yet little factual basis has emerged to warrant the choice of one hypothesized function over another. Empirical evidence has been marshalled in support of each of the contemporary theories. Many are not mutually exclusive but differ primarily in emphasis, each illuminating a different aspect of the whole process. Problem-solving, information-processing, and ego-consolidation theories shade into one another and often appear to be discussing similar or identical processes, despite the use of different languages.

There are a few dimensions upon which some theories of dream function differ substantively from others. One such dimension is that of drive discharge as opposed to balance or dynamic equilibrium as an underlying model. Most drive-discharge approaches, for which Freud's theory is the prototype, tacitly conceive the organism as a relatively static, mechanically closed system that is acted upon from within and without and that moves primarily to rid itself of stimulation in order to return to its original state. Most balance theories, for which Jung's is the psychological prototype, view the organism as a dynamic, biologically active, open system that seeks and interacts with stimulation in such a way as to take in, or integrate, the stimulus in order to arrive at a new and different state of balance. It is noteworthy that Freud, and also Dement (1968) who has been one of the primary contemporary proponents of drive discharge, initially approached the dream via abnormal psychological states. If we begin instead with the normal condition, it is more natural to think, as Fisher (1965) suggests, of a balance mechanism maintaining a steady state of dreaming, with imbalances associated with disturbance.

A second respect in which dream theories clearly differ is in scope. It is a question of aesthetic preference whether one chooses to work with segments or wholes; whether one prefers the sense of certainty that can come from focusing on a limited area or would rather tolerate the ambiguity involved in working with more inclusive concepts. For those who prefer to think on the grand scale, Jung's principles offer a virtually untapped means of encompassing a great deal of existing data, not only on dreams but also in other areas of symbolic expression like fantasy, play, drama, art, mythology, and religion.

A third real difference among theories is the extent to which subjective evidence and nonrational processes are admissable as data. With the exception of the few who have looked at the creative function of the dream, contemporary theorists have stayed very close to a concept of function that derives from biological adaptation to the outside world. The conceptual leap required to think in terms of inner necessity and subjective function does not come easily. Questions like "Where is it leading, psychologically?" and "What is its psychological meaning or value?" have often been confused with metaphysical questions about a prime mover with a purpose outside human psychology. This confusion is unnecessary and leads away from many of the central problems of psychology, the points at which man differs from other animals. Because scientific psychology, which must exclude nonrational *methods*, tends unnecessarily also to exclude nonrational *data*, we split into radical positivists on the one side and radical humanists on the other, each leaving out an essential aspect of a complete psychology. Jung has made room for nonrational data in his nevertheless rational theory of dream function.

Within certain very broad limits, our theories reflect our values which in turn affect the values of those who turn to us as experts. American psychology has grown up in an atmosphere of materialistic and utilitarian values which has had no small effect on our preference for materialistic and utilitarian theories. If we believe that control of the environment is the only important human psychological value, then we will consider functional only processes that promote environmental control. If we value artistic creation, harmony with our neighbors, or a sense of meaning in life as important goals in and of themselves, then psychological processes leading to these ends will be considered functional. Dream psychology is one area in which it is difficult for long to close the door to the latter, specifically human functions.

REFERENCES

ADLER, A. On the interpretation of dreams. *International Journal of Individual Psychology*, 1936, 2, 3–16.

BREGER, L. Function of dreams. *Journal of Abnormal Psychology Monograph*, 1967, **72**(5, Pt. 2, Whole No. 641).

BREGER, L. Children's dreams and personality development. In J. Fisher & L. Breger (Eds.), *The meaning of dreams*. California Mental Health Research Symposium, 1969, No. 3, 64–100.

DEMENT, W. Discussion of F. Snyder, Toward an evolutionary theory of dreaming. *American Journal of Psychiatry*, 1966, **123**, 136–142.

DEMENT, W. The biological role of REM sleep. In A. Kales (Ed.), *Sleep: Physiology and pathology*. Philadelphia: Lippincott, 1968.

ELLENBERGER, H. *The discovery of the unconscious: The history and evolution of dynamic psychiatry*. New York: Basic Books, 1970.

Eriksen, E. The dream specimen of psychoanalysis. *Journal of the American Psychoanalytic Association*, 1954, **2**, 5–55.

Fisher, C. Psychoanalytic implications of recent research on sleep and dreaming. *Journal of the American Psychoanalytic Association*, 1965, **13**, 197–303.

Fiss, H. The need to complete one's dreams. In J. Fisher & L. Breger (Eds.), *The meaning of dreams.* California Mental Health Research Symposium, 1969, No. 3, 38–63.

Foulkes, D. Theories of dream formation and recent studies of sleep consciousness. *Psychological Bulletin*, 1964, **62**, 236–247.

French, T. *The integration of behavior.* Chicago: University of Chicago Press, 1952.

French, T., & Fromm E. *Dream interpretation.* New York: Basic Books, 1964.

Freud, S. *The interpretation of dreams.* London: Hogarth Press, 1953. (Originally published: 1900.)

Greenberg, R. Dreaming and memory. In E. Hartmann (Ed.), *Sleep and dreaming.* Boston: Little, Brown, 1970.

Griffiths, R. *A study of imagination in early childhood.* London: Kegan Paul, 1935.

Hawkins, D. R. A review of psychoanalytic dream theory in the light of recent psychophysiological studies of sleep and dreaming. *British Journal of Medical Psychology*, 1966, **39**, 85–104.

Hollander, M. Is the wish to sleep a universal motive for dreaming? *Journal of the American Psychoanalytic Association*, 1962, **10**, 323–328.

Jones, R. M. *Ego synthesis in dreams.* Cambridge, Mass.: Schenkman, 1962.

Jones, R. M. The psychoanalytic theory of dreaming: 1968. *Journal of Nervous and Mental Disease*, 1968, **147**, 587–604.

Jones, R. M. *The new psychology of dreaming.* New York: Grune & Stratton, 1970.

Jung, C. G. *Psychology and alchemy.* New York: Pantheon, 1963. (Originally published: 1944.)

Jung, C. G. Instinct and the unconscious. In, *The structure and dynamics of the psyche.* New York: Pantheon, 1960. (Originally published: 1919) (a).

Jung, C. G. On the nature of dreams. In, *The structure and dynamics of the psyche.* New York: Pantheon, 1960. (Originally published: 1945) (b).

Jung, C. G. On the relation of analytical psychology to poetry. In, *The spirit in man, art, and literature.* New York: Pantheon, 1966. (Originally published: 1922.) (a)

Jung, C. G. Psychology and literature. In, *The spirit in man, art, and literature.* New York: Pantheon, 1966. (Originally published: 1930) (b).

Klein, G. Peremptory ideation. In R. Jessor & S. Feshbach (Eds.), *Cognition, personality and clinical psychology.* San Francisco: Jossey-Bass, 1967.

Krippner, S., & Hughes, W. Genius at work. *Psychology Today*, 1970, **4**, No. 1, 40–43.

Lerner, B. Dream function reconsidered. *Journal of Abnormal Psychology*, 1967, **72**, 85–100.

Maeder, A. Über die Funktion des Traumes. *Jahrbuch für psychoanalytische und psychopathologische Forschungen*, 1913, **5**, 647–686.

Pearlman, C. A., Jr. The adaptive function of dreaming. In E. Hartmann (Ed.), *Sleep and dreaming.* Boston: Little, Brown, 1970.

Piaget, J. *Play, dreams, and imitation in childhood.* London: Heinemann, 1951.

Shapiro, A. Dreaming and the physiology of sleep. *Experimental Neurology*, 1967, **4** (Suppl. 4), 56–80.

Shapiro, A. A theoretical model for empirical data on the psychophysiology of sleep and dreaming. *Psychophysiology*, 1968, **4**, 375–376.

Trosman, H. Dream research and the psychoanalytic theory of dreams. *Archives of General Psychiatry*, 1963, **9**, 9–18.

Ullman, M. The adaptive significance of the dream. *Journal of Nervous and Mental Disease*, 1959, **129**, 144–149.

Ullman, M. The social roots of the dream. *American Journal of Psychoanalysis*, 1960, **20**, 180–196.

Ullman, M. Dreaming, life-style, and physiology: A comment on Adler's view of the dream. *Journal of Individual Psychology*, 1962, **18**, 18–25.

Whitmont, E. The complex of identity: The ego. In, *The symbolic quest.* New York: C. G. Jung Foundation, 1969.

Witkin, H. Presleep experiences and dreams. *California Mental Health Research Symposium*, 1969, No. 3, 1–37.

(Received September 11, 1972)

OF EDUCATIONAL AND PSYCHOLOGICAL
LIBRARY